

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORLED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)		A2	(11) International Publication Number: WO 99/24578
			(43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: PCT/IB98/01665		(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).	
(22) International Filing Date: 9 October 1998 (09.10.98)			
(30) Priority Data:		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
9723516.2 6 November 1997 (06.11.97) GB 9724190.5 14 November 1997 (14.11.97) GB 9724386.9 18 November 1997 (18.11.97) GB 9725158.1 27 November 1997 (27.11.97) GB 9726147.3 10 December 1997 (10.12.97) GB 9800759.4 14 January 1998 (14.01.98) GB 9819016.8 1 September 1998 (01.09.98) GB			
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).		Published	
(72) Inventors; and		Without international search report and to be republished upon receipt of that report.	
(75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT); RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).			
(54) Title: NEISSERIAL ANTIGENS			
(57) Abstract			
<p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines, supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (eg. 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (eg. 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 5 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide
10 sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise
15 at least n consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, n is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences
20 complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as
25 those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines,
5 for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the
10 presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient
15 a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

20 A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

25 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

5 General

- The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and*
10 *ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene*
15 *Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).
- 20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

- 25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a
5 Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous
10 unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7
15 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the
20 Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes
25 a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems;
30 for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnsteil et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicon systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

10 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.

20 ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion
5 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

10 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,
15 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from
20 wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

25 Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).
30

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg. Summers and Smith supra.*

5 The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as
10 chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

15 In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

20 There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in
25 Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by
30 gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins: in: Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Rept.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g*-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and Chey [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907],

Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*;

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent
5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach and Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

10 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised
15 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably
20 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection
25 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating
30 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described
5 above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of
10 the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then
15 cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes
20 are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A,
25 and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of
30 this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of
5 either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or
10 antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine
15 experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such
20 as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus
25 particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack
30 Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

- 10 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

- Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

- Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc;
- 30 (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

5 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

10 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

20 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; 5 WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 10 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors

15 employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654.

20 Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in 25 which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted 30 terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317;
- 10 Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805;
- 15 Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240;
- 20 Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylogach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245;
- 25 Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre
- 30 (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

- 10 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

- One example are polypeptides which include, without limitation: asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

- 5 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can
10 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful
15 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the
20 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

“Hybridization” refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

- Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed
5 after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid
10 probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will
15 encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some
20 variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe
25 sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as
30 temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

- 5 Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.*

10 backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

- 15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with
- 20 duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern

25 blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid

30 support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl. 11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psорт.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

- 10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).

- 15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one CHCl_3 /isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

- Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *XhoI* restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamHI-XhoI* or *EcoRI-XhoI*), and pET21b+ (using either *NdeI-XhoI* or *NheI-XhoI*).

5'-end primer tail: CGCGGATCCCATATG (*BamHI-NdeI*)

5 CGCGGATCCGCTAGC (*BamHI-NheI*)

CCGGAATTCTAGCTAGC (*EcoRI-NheI*)

3'-end primer tail: CCCGCTCGAG (*XhoI*)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF;

10 the same 3' *XhoI* primer was used as before:

5'-end primer tail: GGAATTCCATATGGCCATGG (*NdeI*)

5'-end primer tail: CGGGATCC (*BamHI*)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI*-

15 *BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (*NheI*)

3'-end primer tail: CGGGATCC (*BamHI*)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

25 The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD₂₆₀ was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

10 D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

- Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or
- 25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the
10 vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation
10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM
15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21µl of each
20 fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-
25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

10 M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

10 **O) FACScan bacteria Binding Assay procedure.**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

15 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 A. GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAQGN AAAQYNLGAM
51  YXQRTVRVRD DAEAVRWYRQ PAEQGLAQAG YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVQAQ
151 AQNNLGVMYA ERXRVQRD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
51 YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAO YNLGWMYANG RGVRRDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151 AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGGTTAT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
51 YAERRGVRRD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *

```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

		10	20	30	40	50	60
20	orf37.pep	MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRRRD					
	orf37a	MKQTVKWLAAALIALGLNQAVWADDVSDFR ENLQAAAQGNAAAQNNLGVMYAERRGVRRD					
		10	20	30	40	50	60
25		70	80	90	100	110	120
	orf37.pep	DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG					
		:	: :				
	orf37a	RALAQEWLGKACQNGYQDSCDNDQRLKAGYX					
		70	80	90			

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
151 TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
35 201 GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251 GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGGCGCTCGCT
301 CAACAATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAAA ACAGCTGCGA
351 CAATGACCAA CGCCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
40 51 YENGQGVRRD YVQAVQWYRK ASEQGDAAQAO YNLGLMYDGR RGVRRDLALA
101 QQWLKGACQN GDQNSCDNDQ RLKAGY*

```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

45	orf37.pep	MKQTVXMLAAALIALGLNRPVWXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRRRD	60
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSDFR ENLQAAEQGNAAAQFNLGVMYENGQGVRRD	60
50			
	orf37.pep	DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG	120
		: : : :	: :
	orf37ng	YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRQDLALAQQWLKGACQNGDQNSCDNDQ	120
	orf37.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRQD	168
55	orf37ng	RLKAGY	126

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

		10	20	30	40	50	60
5	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSD	FRENLQAAAQGNAAAQYNLGAMYYKGRGVR	RD			
	orf37ng	MKQTVKWLAAALIALGLNQAVWAGDVSD	FRENLQAAEQGNAAAQFNLGVMYENGQGV	RQD			
		10	20	30	40	50	60
10	orf37-1.pep	DAEAVRWYRQAAEQGLAQAYNLGWMYANG	RGVRQDDTEAVRWYRQAAAQGVVQAQYNL	G			
	orf37ng	YVQAVQWYRKASEQGDAQAYNLGLMYDGR	GVQRQD-----				
		70	80	90			
15	orf37-1.pep	VIYAEGRGVRQDDVEAVRWFRQAAAQGV	AQAQNNLGVMYAERRGVQRDRALAQEWFG	KAC			
	orf37ng	-----	LALAQQWLKAC				
			100				
20	orf37-1.pep		190	199			
	orf37ng		110	120			

- 25 Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

- 35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

40	TTCCGCGA	CATCGGCGGT	TTGAAGGTCA	ATGCCCCCGT	CAAATCCGCA
	GGCGTATTGG	TCGGGCGCGT	CGGCGCTATC	GGACTTGACC	CGAAATCCTA
	TCAGGCGAGG	GTGCGCCTCG	ATTTGGACGG	CAAGTATCAG	TTCAGCAGCG
	ACGTTTCCGC	GCAAATCCTG	ACTTCSGGAC	TTTGGGCGA	GCAGTACATC
	GGGCTGCAGC	AGGGCGGCGA	CACGGAAC	CTTGCTGCCG	GCGACACCAT
	CTCCGTAACC	AGTTCTGCAA	TGTTCTGGA	AAACCTTATC	GGCAAATCA
45	TGACGAGTTT	TGCCGAGAAA	AATGCCGACG	GCGGCAATGC	GGAAAAAGCC
	GCCGAATAA				

This corresponds to the amino acid sequence <SEQ ID 10>:

1	FGDIGGLKVN	APVKSAGVLV	GRVGAIGLDP	KSQYARVRLD	LDGKYQFSSD
51	VSAQILTSGL	LGEQYIGLQQ	GGDTENLAAG	DTISVTSSAM	VLENLIGKFM

101 TSFAEKNADG GNAEKAAE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
ybrd.h LGIGALVFLGLRVANVQGF AETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
          10      20      30

10     80      90      100     110     120     130
ybrd.h KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
N.m      KSYQARVRLDLGKY-QFSSDVSAQILTSGLLGEQYIGLQOG---GDTENLAAGDTISVT
          40      50      60      70      80

15     140     150     160
ybrd.h TSAMVLEDLIGQFL--YGSKSDGNEKSESTEQ
N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
          90      100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25     20      30      40      50      60      70
ybrd      GAAVAFLAFRVAGGAFFGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
          10      20      30

30     80      90      100     110     120     130
ybrd      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQOGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQOGGDTENLAAGDTISVTSSAM
          40      50      60      70      80      90

35     140     150     160
ybrd      VLENLIGKFMTSFAEKNADGGNAEKAAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAAEX
          100     110     120

```

The complete ybrd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1  ..ATTTGATAT ACCTCATCCG CAAGAATCTA GGTTCCGCCG TCTTCTTCTT
      51  TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAATG GTCAAATTCC
      101  GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
      151  GAACGCCTGA CACCGTTCGG CAAAAAACTG CGTGCCGcCA GTwTGGACGA
      201  ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
      251  CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
      301  CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```


-64-

5
10

```

351 GCGCAACGcg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451 AAAAAAGTAT TAATCAAGGA AGGGATTTC GCACAGGGCG AACA.aCCAT
501 GCCCCCTTTC ACAGGAAAAAC GCAAACTCGC CGTCGTCTGGT GCGGGCGGAC
551 ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
601 ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651 CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701 ACGTCGCGGT CGCCGTGCGC AACAAACCGCA TCCGCCGCCA AATCGCCGAA
751 AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
801 GACCGTCTCG CTTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851 AAGCGGTCG..

```

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

15

```

1  ..ILIYLIRKLNLS GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
51  ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQW
101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTV
151 KKVLIKEGIS AQGEXTMPFF TGRKRLAVVG AGGHGKVVAD LAAALGRYRE
201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20
25
30
35
40

```

1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51  ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101 AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
151 GGAACACCTT TTAATATGGT CAAATTCCTG TCCATGCGCG ACGCGTTGA
201 TTCAGACGGC ATTCGCTGTC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251 AAAAATCGCG TGCCGCCAGT TTGGACGAAC TGCTGAATT ATGGAATATC
301 TTAAGAGCGG AGATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
451 GAAAAATTCG CCTCGCATGT TTGTATATAC GACCACTTCA GCCTGTGCCT
501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551 GGATTTCCCG ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAACCGC
601 AAATCGCCG TCGTCGGTGC GGGCGGACAC GGAAGTTCG TTGCCGACCT
651 TGCCGCCGCA CTCGCCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
751 GAAAAACAGT TATCGCCGGA ACAATACGAC GTCGCCGTCG CCGTCCGCAA
801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
901 GTCGACAAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001 ACTGCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCTG
1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGACC TCGACAGCAT AA

```

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

50

```

1  MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSPV FFFQERPGKD
51  GKPFKMVKFR SMRDALDSG IPLPDGERLT PFGKKLRAAS LDELPELWNI
101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
151 EKFACDVWYI DHFSLCLDIK ILLLVKKVL IKEGISAQGE ATMPPTGKR
201 KLAVVGAGGH GKVVDLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLL
251 ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
301 VGQGSVVMK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPAHLS
351 GNTHIGESW IGTGACSRQQ IIRIGSRATIG AGAVVVRDVS DGMTVAGNPA
401 KPLPRKNPET STA*

```

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

-65-

orf3.pep ILIIYLIRKNLGSPVFFQERPGKDGKPFKMVKFR
 orf3a MSKFFFKRLFDIVASASGLIFLSPVFLILIIYLIRKNLGSPVFFQERPGKDGKPFKMVKFR
 10 10 20 30 40 50 60

orf3.pep SMRDGLYSDGIPDPGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL
 orf3a SMHDAALSDGILLDPGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
 10 40 50 60 70 80 90 100 110 120

orf3.pep YDNFQNRHRHEMKPGITGWAQVNGRNALSWDEKFCADVWYIDHFSCLCDIKILLTVKKVL
 orf3a YDNFQNRHRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSCLCDIKILLTVKKVL
 15 100 110 120 130 140 150 160 170 180

orf3.pep IKEGISAQGEEXTMPFFTGKRKLAVVGAGGHGKVVAADLAAALGRYREIVFLDDRAQGSVNG
 orf3a IKEGISAQGEATMPFFTGKRKLAVVGAGGHGKVVAELAAALGTGEIVFLDDRQGSVNG
 20 160 170 180 190 200 210 220 230 240

orf3.pep FSVIGTTLLENLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
 orf3a FSVIGTTLLENLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLVHPDSTVSPSAT
 25 220 230 240 250 260 270 280 290 300

orf3.pep VGQGSVVMKAV
 orf3a VGQGSVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLGSGNTRIGESW
 30 280 310 320 330 340 350 360

35 The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
 51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
 101 AGAATCTGGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
 151 GGAAACCTT TTAATATGGT CAAATTCGT TCCATGCACG ACGCGCTTGA
 40 201 TTCAGACGGC ATCTGCTGTC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 251 AAAAAGTGGC TGCCGCCAGT TTGGACGAAC TGCCCGAACT GTGGAACGTC
 301 CTCAAAGGCG ACATGAGCCT GGTCCGCCCC CGCCCGCTGC TGATGCAATA
 351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCGG
 401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 45 451 GAACGCTTCG CATGCCACAT CTGGTATATC GACCACTTCA GCCTGTGCCT
 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAAGAAG
 551 GGATTTCGCG ACAGGCGGAA GCCACCATGC CCCCTTCAC AGGAAAACGC
 601 AAAGTTGCGG TCGTCGGTGC GGGCGGACAC GGCAAGTTCG TTGCCGAGCT
 651 TGCCGCGGCA CTCGGCACAT ACGGCGAAAT CGTTTTCTCG GACGACCGCG
 50 701 TCCAAGGCAG CGTCAACGGC TTCCCGTCA TCGGCACGAC GCTGCTGCTT
 751 GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCGCCGTCG CCGTCGGCAA
 801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCGCGC CTCGGCTTCG
 851 CCCTGCCCGT CCTGATTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA
 901 GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCGTCGTAC AGGCTGACAG
 55 951 CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCAGC
 1001 ATTGCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTCG
 1051 GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG
 1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
 1151 TCGTCGTGCG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAACCCGGCA
 60 1201 AAACCATTGG CAGGCAAAAA TACCGAGACC CTGCGGTCGT AA

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

1 MSKFFFKRLFD IVASASGLIF LSPVFLILII LIRKNLGSPV FFFQERPGKD
 51 GKPFKMVKFR SMHDAALSDG ILLPDGERLT PFGKKLRAAS LDELPELWNV
 101 LKGDMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
 151 ERFACDIWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPFFTGKR
 201 KLAVVGAGGH GKVVAELAAA LGTYGEIVFL DDRVQGSVNG FSVIGTLLL
 65 251 ENSLSPEQFD IAVAVGNNRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT

301 VGQGGVMAK AVVQADSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHL S
 351 GNTRIGEE SW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KPLAGKNTET LRS*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFQERPGKDGKPFKMKVKFR				
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNL	GSPVFFQERPGKDGKPFKMKVKFR				
10		10	20	30	40	50	60
	orf3a.pep	SMHDALDSGILLPDGERLTPFGKKLRAASLDELPELWNV	LKGDSL VGPRPLLMQYLP				
	orf3-1	SMRDALDSGIPLPDGERLTPFGKKLRAASLDELPELWNL	KGEMSL VGPRPLLMQYLP				
15		70	80	90	100	110	120
	orf3a.pep	YDNFQNRHMKPGITGWAQVNGRNL	SWDERFACDIWYIDHFSCLDIKILLTVKKVL				
	orf3-1	YDNFQNRHMKPGITGWAQVNGRNL	SWDEKFCADVWYIDHFSCLDIKILLTVKKVL				
20		130	140	150	160	170	180
	orf3a.pep	IKEGISAQGEATMPPTGKRKLAVV	GAGGHGKVVAELAAALGTYGEIVFLDDR	VQGSVNG			
	orf3-1	IKEGISAQGEATMPPTGKRKLAVV	GAGGHGKVVAELAAALGRYREIVFLDDR	AQGSVNG			
25		190	200	210	220	230	240
	orf3a.pep	FPVIGTTLLENSLSPEQFDIAVAVGNNRIR	QIAEKAAALGFALPVLHPDSTVSPSAT				
	orf3-1	FSVIGTTLLENSLSPEQYDVAVAVGNNRIR	QIAEKAAALGFALPVLVHPDATVSPSAT				
30		250	260	270	280	290	300
	orf3a.pep	VGQGGVMAKAVVQADSVLKDGIVIVNTAATVDH	DCLLDAFVHISPGAHL	SGNTRIGEE SW			
	orf3-1	VGQGSVMAKAVVQAGSVLKDGIVIVNTAATVDH	DCLLNAFVHISPGAHL	SGNTHIGEE SW			
35		310	320	330	340	350	360
	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLAGKNTETLRSX				
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX				
40		370	380	390	400	410	
	orf3a.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLAGKNTETLRSX				
	orf3-1	IGTGACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX				
45		370	380	390	400	410	

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKNL	GSPVFFQERPGKDGKPFKMKVKFR	SMRDGLYSDGIPLPDGERLTPFGKKLRA	62
	yvfc	27	I++R +GSPVFF Q RPG GKPF + KFR+M D S G LPD RLT G+ +R			
			I+VRLKIGSPVFFKQVRPGLHGKPF	TLTKFRTMTDERDSKGNLLPDEVRLTKTGR	LIRK	86
55	ORF3	63	ASXDELPELWNLKGEMSLVGPRPLLMQYLP	YDNFQNRHMKPGITGWAQVNGRNL	122	
	yvfc	87	LSIDELPQLNLVNLGDSL VGPRPLLMQYLP	YDKQARRHEVKPGITGWAQINGRNAIS	146	
60	ORF3	123	WDEKFCADVWYIDHFSCLDXXXXXXXXXXXX	XEGISAQGEATMPPTG	172	
	yvfc	147	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLV	SEGIQQTNHVTAERFTG	196	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

5	orf3	ILIIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKKNKGSPPVFFIRERPGKDGKPFKMVKFR	60
10	orf3	SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLMQYLPL	94
	orf3ng	SMRDALDSGDIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLMQYLPL	120
15	orf3	YDNFQNRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLDIKILLTVKKVL	154
	orf3ng	YNKFQNRHEMKPGITGWAQVNGRNALSWDEKFSQDVWYTDNFSFWLDMKILFLTVKKVL	180
20	orf3	IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAAELAAALGTYGEIVFLDDRTQGSVNG	240
25	orf3	FSVIGTTLLENLSLSPEQYDVAVAVGNRRIRRIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTTLLENLSLSPEQFDITVAVGNRRIRRIQITENAAALGFKLPVLIHPDATVSPSAI	300
30	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEEER	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
51	GCTGATTGTC	CTGTCGCCCG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
101	AAAACTTAGG	TTCCGCCGTC	TTCTTCattc	GGGAACGCCc	cgGAAAGGAc
151	ggaaaacCTT	TTAAATGGT	CAAATTCCGT	TCCatgcgcg	acgcgcttGA
201	TTCAGACGGG	ATTCCGCTGC	CCGATAGCGA	ACGCTGACC	GATTTCGGCA
251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTGAATT	ATGGAATGTC
301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTT	TGATGCAGTA
351	TCTGCCGCTT	TACAACAAAT	TTCAAAACCG	CGCCACGAA	ATGAAACCGG
401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCTGGGAC
451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACAATTTCA	GCTTTTGGCT
501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
551	GCAATTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
651	TGCCGCGCGA	CTCGGCACAT	ACGCGGAAAT	CGTTTTTCTG	GACGACCGCA
701	CCCAAGGCAG	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCACCGTCG	CCGTCGGCAA
801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
851	AATGCCCCGT	TCTGATTTCAT	CCCACGCGCA	CCGTCTCGCC	TTCTGCAATA
901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCGTAC	AGGCCGGCAG
951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTGATCACG
1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCTG
1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgcccGT	GCAGGGgcGG
1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCCGGGG	CAACCCGGCA
1201	AAGCCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKNLGSPV	FFIRERPGKD
51	GKPFKMVKFR	SMRDALDSG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ	VNGRNALSWD
151	EKFSQDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
201	KLAVIGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTTLLL
251	ENSLSPEQFD	ITVAVGNNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
301	IGQGSVVMK	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLDAF	VHISPGAHL
351	GNTRIGEEER	IGTGACSRQQ	TTVGSGVTAG	AGAVIVCDIP	DGMTVAGNPA
401	KPLTGKNPKT	GTA*			

-68-

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKNLGSPVFFFQERPGKDGKPFKMKVKFR					
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKNLGSPVFFIRERPGKDGKPFKMKVKFR					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf3-1.pep	SMRDALDSDGIPLPDGERLTDFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
	orf3ng	SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf3-1.pep	YDNFQNRHMKPGITGWAQVNGRNLASWDEKFCADVWYIDHFSCLDIKILLTVKKVL					
	orf3ng	YNKFQNRHMKPGITGWAQVNGRNLASWDEKFCADVWYIDNFSEWLDKILFLTVKKVL					
		130	140	150	160	170	180
		190	200	210	220	230	240
20	orf3-1.pep	IKEGISAQGEATMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG					
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAAELAAALGTGEIVFLDDRTQGSVNG					
		190	200	210	220	230	240
25		250	260	270	280	290	300
	orf3-1.pep	FSVIGTTLLENSLSPEQYDVAVAVGNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
30	orf3ng	FPVIGTTLLENSLSPEQFDITVAVGNRIRRQITENAAALGFKLPVLIHPDATVSPSAI					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf3-1.pep	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEEESW					
35	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEEESR					
		310	320	330	340	350	360
		370	380	390	400	410	
40	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVSDGMTVAGNPAKPLPRKNPETSTAX					
	orf3ng	IGTGACSRQQTTVSGSVTAGAGAVIVCDIPDGMTVAGNPAKPLTKGNPKTGTAX					
		370	380	390	400	410	

In addition, ORF3ng shows significant homology with a hypothetical protein from *B.subtilis*:

45	gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis] Length = 202
	Score = 235 bits (594), Expect = 3e-61
50	Identities = 114/195 (58%), Positives = 142/195 (72%)
	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLRKNLGSPVFFIRERPGKDGKPFKMKVKFRSMRD 64
	+KRLFD+ A+ L S + L I ++R +GSPVFF + RPG GKPF + KFR+M D
	Subjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVVRLLKIGSPVFFKQVRPGLHGKPFFTLYKFRTMTD 62
55	Query: 65 ALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPLYNKF 124
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPY +
	Subjct: 63 ERDSKGNLLPDEVRLTKTGRLLKLSIDELPQLLNVLKGDLSLVGPRPLLMYDLYPLYTEK 122
60	Query: 125 QNRHMKPGITGWAQVNGRNLASWDEKFCADVWYIDNFSEWLDKILFLTVKKVLIKEG 184
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Subjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEG 182
	Query: 185 ISAQGEATMPPFAGN 199
	I T F G+
65	Subjct: 183 IQQTNHVTAERFTGS 197

The hypothetical product of *yvf* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGCT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTGCGAC GCGCCGACAA CCGCCGCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCGC CGTTTCTGCA
15 401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTS RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCC
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
25 151 AAAGTCCTCG ATTTTCTCGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCGTTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAGA GTTCCGCGAA
30 451 CAGCGCAACC ATATGGCGAT TGTATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCCGGCAG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCGTGGTC
35 701 ATTCAAGAGT TGGGACATCT GCCCGTGGC GCGGAAAAAG TCCTTATCGG
751 CGGTTTGAG TTACCCGTC CACGCGCCGA CAACCGCCGC CTGCATACGC
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTAGGA
851 TGACGGTACG GCGGTTTTCT GTTCAATCC GCCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLRLLE
51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKMFNP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPHSRVGTG ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCC
51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
50 151 AAAGTCCTCG ATTTTCTGTA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCGTTCAT CGGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CTGCGGTCT

```

This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

```

1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLRLRLE
51 KVLDFDSFLE RDAMITRSM NVLKENDSIE RITAYVIDTA HSRFPVJGED
101 KDEVGLILHA KDLLKYMENP QEFHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIQIVG DIEDEFDEDE SADNIHAVSA
201 ERWRIHAATE IEDINAFFGT EYSSEADTI GGXGHSIGIT PARARRKSYX
251 RRRAXHXRRX XQPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT
301 *

```

25

35

40

45

50

55

60

	1	MDGAQPKTNF	FERLIARLAR	EPDSAEDVLN	LLRQAHEQEV	FDADTLTRLE
	51	KVLDFAEVL	RDAMITSRM	NVLKENDSIE	PATAYVIDTA	HSRFPVIGED
	101	KDEVLGILHA	KDLLKMFNP	EQFHLKSVLR	RITAVPEGKS	LTALLKEGED
10	151	QRNHMAIVID	EYGGTSGLVT	FEDIIEQIVG	DIEDEFEDE	SADDIHSVSA
	201	ERWRTHAATE	IEDINAFGT	EYGSEADCI	RRHGSGT	PARARRKSPY
	251	RRFAVHRRPR	ROPPPAHADG	DPREVSRADP	HLRFCTV*	

15	1	ATGGACGGCG	CACAACCGAA	AACAAATTTT	TTTGAACGCC	TGATTGCCCG
	51	ACTCGCCCGC	GAACCCGATT	CCGCCGAAGA	CGTATTAAAC	CTGCTTCGGC
	101	AGGCGCAGCA	ACTGGAAGTT	TTTGATGCCG	ACACACTGAC	CCGGCTGGAA
	151	AAAGTATTGG	ACTTTCCCGA	GCTGGAAGTG	CGCGATGCGA	TGATTACCGC
20	201	CAGCCGCATG	AACGTATTGA	AAGAAAACGA	CAGCATCGAA	CGCATCACCG
	251	CCTACGTCAT	CGATACCGCC	CATTTCGCGCT	TCCCGCTCAT	CGGCGAAGAC
	301	AAAGACGAAG	TTTTGGGCAT	TTTGCACGCC	AAAGACCTGC	TCAAATATAT
	351	GTTCACCCCC	GAGCAGTTCC	ACCTGAAATC	CGCTTTCGCG	CCTGCGGTTT
25	401	TCGTGCCCGA	AGGCAAATCT	TTGACCGCCC	TTTTAAAAGA	GTTCCGCGAA
	451	CAGCGCAACC	ATATGGCAAT	CGTCATCGAC	GAATCTCGCG	GCACGTCGGA
	501	TTTGGTCAAC	TTTGAAGACA	TCATCGAGCA	AATCGTCGGT	GACATCGGAAG
	551	ACGAGTTTGA	CGAAGACGAA	AGCGccgacg	acatCCACTC	cgTTTccgCC
30	601	GAACGCTGGC	GCATCCacgc	ggctaCCGAA	ATCGAAGaca	TCAACGCCTT
	651	TTTCGGTACG	GAatacggca	gcgaagaagc	cgacaccatc	cggcggctTG
	701	GTCATTACAG	AATTGGGACA	CCTGCCCGTG	CGCGGCGAAA	AAGTCTTTAT
	751	cggcgGTTTG	Cagttcacgc	tCGCCCGCGC	CGACAACCCG	CGCCTGCACA
30	801	CGCTGATGGC	GACCCGCGTG	AAGTAAGCAG	AGCCTGCCcg	AccgcggttT
	851	CTGCacAGTT	TAGGatgACG	gtaCGGTCGT	TTTCTGTTTC	AATCCGCCCC
	901	ATCCGCCAAA	CATAA			

35	1	MDGAQPKTNF	FERLIARLAR	EPDSAEDVLN	LLRQAHEQEV	FDADTLTRLE
	51	KVLDFAELEV	RDAMLTRSRM	NVLKENDSIE	RITAYVIDTA	HSRFPVIGED
	101	KDEVLGILHA	KDLLKYMFPN	EQFHLKSVLR	PAVFVPEGKS	LTALLKEFRE
	151	QRNHMAIVID	EYGGTSGLVT	FEDIEQIVG	DEIEDEFEDE	SADDIHSVSA
	201	ERWRIHAATE	IEDINAFFGT	EYGSEADTI	RRLGHSIGIT	PARARRKSPY
40	251	RRFAVHRRPR	RQPPPAHADG	DPREVSRACP	TAVSAQFRMT	VRSFSVSIRP
	301	TROT*				

	orf5	NHMAIVIDEYGGTSGLVT FEDII EQIVGEI	30
45	orf5ng	FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVT FEDII EQIVGDI	182
	orf5	EDEFDEDD SADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTS	90
50	orf5ng	EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSSEADTIRRLGHSGIGTPA	242
	orf5	RARRKSPYRRFAVHRRTRRQPPAYADGDPREVSX---RRFCTV	131
	orf5ng	RARRKSPYRRFAVHRRP RRQPPAHADGDPREVS RACPHRRFCTV	287

10 20 30 40 50 60
orf5nq-1.pep MDGAQPKTNFFERLIARLAREPDSAEVDVLLNLRQAEQEVFDADTLTRLEKVLDFAELEV

Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae*

ORF5 2 HMAIVIDEYGGTSGGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDWTIRIHAATEIED 61
HMAIV+DE+G SGLVT EDI+EQIVG+IEDEFDE++ AD I +S T+ + A T+I+D
TlyC 166 HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5 62 INTFFGTEYSIEEADTI 78
N F T++ EE DTI
TlyC 225 FNAQFNITDFDDEEVDTI 241

SCORES Init1: 301 Initn: 419 Opt: 668
Smith-Waterman score: 668; 45.9% identity in 242 aa overlap

[illegible]

```

                130      140      150      160      170      180
5  orf5ng-1.pep  170      180      190      200      210      220
    VTIEDIIEQIVGDIEDEFDEDESADDIHSVAERWRIHAATEIEDINAFGTEYGSEED
    ||:||||:|||||||:| || |::| : : ::| |:|:|:| | |:|:: :||:|
    tlyc_haein   VTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEV
                190      200      210      220      230
10 orf5ng-1.pep  230      240      250      260      270      280
    TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPERRQPPPAHADGDPREVSRACTAVSAQF
    || | : :| | :|
    tlyc_haein   TIGGLIMQTFGYLPKRGEIILKNLQFKVTSADSRRLIQLRVTVPDEHLAEMNNVDEKSE
                240      250      260      270      280      290

```

15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

```

20 sp|P77392|YBEX_ECOLI_HYPOTHETICAL_33.3_KD_PROTEIN_IN_CUTE-ASNB_INTERGENIC_REGION
    >gi|1778577 (U82598) similar to H. influenzae [Escherichia coli] >gi|1786879
    (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
    approx. 440 aa protein YTFL_HAEIN SW: P44717 [Escherichia coli] Length = 292

```

Score = 212 bits (533), Expect = 3e-54

Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

```

25 Query: 2   DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
    D   K   F   L+++L   EP + +++L L+R + + ++ D DT   LE V+D +D V
    Sbjct: 10 DTISNKKGFFSLLLSQLFHGEPKRNDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

30 Query: 61  RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
    RD MI RS+M LK N +++   +I++AHSRFPVI EDKD + GIL AKDLL +M +
    Sbjct: 70  RDIMIPRSQMITLKRNTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

35 Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVTFEDIIEQIV 179
    E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
    Sbjct: 130 AEAFSMDKVLQAVVPEKSRVDRMLKEFRSQRVHMAIVIDEFGGVSGGLVTIEDILELIV 189

    Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFGTEYSSEADT 229
    G+IEDE+DE++ D   +S   W + A   IED N   FGT +S EE DT
    Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVDT 238

```

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

45 ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```

    1  ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTTC
    51  GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC

```

-74-

101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
 151 GATGCCTTCA GCGGCAATCC TGAAgGGCAG TTTTTCCTCCG ACAGCTACGA
 201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAgG
 5 251 GCGATGCAAC GCGGCTGAA TGAgGGCATG GGAAAGCAGG CAGGACGGGC
 301 TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
 351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
 401 CAACCGCTTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
 451 ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
 10 501 CGCGACACGC CGTACAACAC CTACACGCGC GCGGTCTGC CGCCAACCCC
 551 GATTGCGCTG CCC..

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

1 MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51 DAFSGNPEQ FFPDSYEIDA GGS DLQIYQT AYKAMQRRLN EAWESRQDGL
 15 101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFNRLKIGM LQTXSVIY
 151 GMGAAYKGKI RKADLRDTP YNTYTRGGLP PTPIALP..

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCGGC
 51 AGCCGTTTTT CCGCGCTGCG TTTTGTTC TAAGGATAAC GGCAGGGCAT
 101 ACCGAATCAA AATTGCCAAA AACCAGGTA TTTCTCGGT CGGCAGGAAA
 20 151 CTGCGCGAAG ACCGCATCGT GTTACGAGG CATGTTTGA CGCGCGCGGC
 201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
 251 CTTGCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGCGCGCAGG
 301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTGCGATAT
 351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
 25 401 GGAGCAATGA AAAACTGATG CGGGAAGTTG CGCCCGATGC CTTACGCGGC
 451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
 501 CAGTGATTG CAGATTAC AAACCGCCTA CAAGGCGATG CAACGCCGCC
 551 TGAATGAGG ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
 601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
 30 651 AGCCGACCGC GACCATGTCT CTTCCGTCTT CGTCAACCGC CTGAAATCG
 701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
 751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
 801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
 851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
 35 901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
 951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

1 MLRKLLKWSA VELTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
 51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQKMRGGR
 40 101 PDSVTVQIIE GSRFSMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
 151 NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLEAWES RQDGLPYKNP
 201 YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYMGMAA
 251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
 301 FVSKMDGTGL SQFSDLTEH NAAVRKYILK K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H. influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

ORF7 1 MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
 + G+ V+ IEG F RK ++ P + K SNE++ A ++ +
 50 yceg 102 LNSGKEVQFNVKWIEGRTFKDWRKDLENAPHLVQTLKDKSNEEIFALLDLDPIDIGQNLK 161
 ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLV 115
 N EG +PD+Y +DL++ + + M++ LN+AW R + LP NPYEMLI+A +V
 yceg 162 NVEGWLYPDTYNTYTPKSTDLLELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV 221
 55 ORF7 116 EKETGHEAXXDHVASVFVNRLLKIGMRLQTXSVIYMGMAAYKGKIRKADLRDTPYNTYT 175
 EKETG VASVF+NRLK M+LQT +VIYMG Y G IRK DL TPYNTY
 yceg 222 EKETGIANERAKVASVFVNRLLKAKMKLQTDPTVIYMGGENYNGNIRKKDLETKTPYNTYV 281

ORF7 176 RGGLPPTPIALP 187
GLPPTPIA+P
yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDLLD
     101 LLNSGKEVQF NVKWIEGKTF KDWRKDLENA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLEL KNVEGWLYPD TYNYPKSTD LELLKRSAER MKKALNKAWN
     201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
    10 251 DPTVIYGMGE NYNGNIRKRD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
     301 VANPEKTDYF YFVADGSGGH KFTRNLNEHN KAVQEYLRWY RSQKNAK

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N.*

15 *meningitidis*:

```

      10      20      30
orf7.pep      MRGGRPDsvTVQIIeGSRfSHMRKVIdATP
      |||||||||||||||||||||||||||||||||||
20  orf7a      AAYVLGVHnRLHTgTYRLPSEvSAWDILQkMRGGRPDsvTVQIIeGSRfSHMRKVIdATP
      70      80      90      100     110     120

      40      50      60      70      80      90
orf7.pep      DIGHDTKGWSNEKLMAEvAPDAfSGNPEGQFFPDsYEIdAGGSdLQIYQTAYKAMQRRLN
      |||||||||||||||||||||||||||||||||||:|||||
25  orf7a      DIEHDTKGWSNEKLMAEvAPDAfSGNPEGQFFPDsYEIdAGGSdLRIYQIAYKAMQRRLN
      130     140     150     160     170     180

      100     110     120     130     140     150
orf7.pep      EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDhVASvFVNRLKIGMRlQTXXSVIY
      |||||||||||||||||||:|||||||||||||||||
30  orf7a      EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDhVASvFVNRLKIGMRlQTDPsVIY
      190     200     210     220     230     240

      160     170     180
orf7.pep      GMGAAYK GKIRKADLRrdTPYNTYTRGGLPPTPIALP
      |||||||||||||||||||||||||||||||||||
35  orf7a      GMGAAYK GKIRKADLRrdTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFvSKM
      250     260     270     280     290     300

40  orf7a      DGTGLSQFSHDLTEHNAAVRKYILKKX
      310     320     330

```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

      1 ATGTTGAGAA AATGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
      51 AGCCGTTTTTC GCCGCGCTGC TTTCGTCCTC TAAAGACAAC GGCAGGGCAT
45  101 ACAGGATTAA AATTGCCAAA AACCAGGTA TTTCGTCGGT CGGCAGGAAA
     151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
     201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
     251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
     301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTGCGGTT TTTCGCATAT
50  351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
     401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTCAGCGGC
     451 AATCCTGAAG GGCAGTTTTT CCCCAGACAG TACGAAATCG ATGCGGGCGG
     501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
55  551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
     601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGCATGTA
     651 AGCCGACCGC GACCATGTCG CTTCCGCTTT CGTCAACCGC CTGAAAATCG
     701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
     751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
     801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCCCGGCA
60  851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
     901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
     951 GACCGAACAC AACGCCGCCG TTCGCAAATA TATTTTGAAA AAATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
 51 LAEDRIVFSR HVLTAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
 101 PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
 151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
 201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYMGAA
 251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPGSEKLY
 301 FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

		10	20	30	40	50	60
orf7a.pep		MLRKLLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK	LAEDRIVFSR
orf7-1		MLRKLLKWSA	VFLTVSAAVF	AALLFVPKDN	GRAYRIKIAK	NQGISSVGRK	LAEDRIVFSR
		10	20	30	40	50	60
orf7a.pep		HVLTAAYVL	GVHNRLHTGT	YRLPSEVSAW	DILQKMRGGR	PDSVTVQIIE	GSRFHMRKV
orf7-1		HVLTAAYVL	GVHNRLHTGT	YRLPSEVSAW	DILQKMRGGR	PDSVTVQIIE	GSRFHMRKV
		70	80	90	100	110	120
orf7a.pep		IDATPDIEHD	TKGWSNEKLM	AEVAPDAFSG	NPEGQFFPDS	YEIDAGGSDL	RIYQIAYKAM
orf7-1		IDATPDIEHD	TKGWSNEKLM	AEVAPDAFSG	NPEGQFFPDS	YEIDAGGSDL	RIYQIAYKAM
		130	140	150	160	170	180
orf7a.pep		QRRLEAWES	RQDGLPYKNP	YEMLIMASLI	EKETGHEADR	DHVASVFVNR	LKIGMRLQTD
orf7-1		QRRLEAWES	RQDGLPYKNP	YEMLIMASLI	EKETGHEADR	DHVASVFVNR	LKIGMRLQTD
		190	200	210	220	230	240
orf7a.pep		PSVIYMGAA	YKGIKIRKAD	LRRDTPYNTY	TRGGLPPTPIA	LPGKAALDAA	AHPGSEKLY
orf7-1		PSVIYMGAA	YKGIKIRKAD	LRRDTPYNTY	TRGGLPPTPIA	LPGKAALDAA	AHPGSEKLY
		250	260	270	280	290	300
orf7a.pep		FVSKMDGTGL	SQFSHDLTEH	NAAVRKYILK	KX		
orf7-1		FVSKMDGTGL	SQFSHDLTEH	NAAVRKYILK	KX		
		310	320	330			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N.*

gonorrhoeae:

50	orf7	MRGGRPDSTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM	AEVAPDAFSGNPEGQ	60
	orf7ng	MRGGRPDSTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLM	AEVAPDAFSGNPEGQ	60
55	orf7	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMAXLVEKETG		120
	orf7ng	FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG		120
60	orf7	HEAXXDHVASVFVNRLLKIGMRLQTXSVIYMGAAAYKGIKIRKADLRRDTPYNTYTRGGGLP		180
	orf7ng	HEADRDHVASVFVNRLLKIGMRLQTDPSVIYMGAAAYKGIKIRKADLRRDTPYNTYTGGLP		180
	orf7	PTPIALP		187

orf7ng || ||||
 PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK 236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

5 1 MRGGRPD SVT VQIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51 DAFSGNPEGQ FFPDSYEIDA GGS DLQIYQT AYKAMQRR LN EAWAGRQDGL
 101 PYKNPYEMLI MASLIEKETG HEADRDHVAS VFNRLKIGM RLQTDPSVIY
 151 GMGAAYKGKI RKADLRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG
 201 EKYLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK*

10 Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

 1 ..taccgaatca AGATTGCCAA AAATCAGGGT ATTCGTCGG TCGGCAGGAA
 51 ACTTGCCgaA GACCGCATCG TGTT CAGCAG GCATGTTT TG ACAGCGGCGG
 101 CCTACGTTTT GGGTGTCAC AACAGGCTGC ATACGGGGAC gTACAGATTG
 151 CCTTCGGAAG TGTCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG
 15 201 GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTGCGGT TTTTCGCATA
 251 TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC
 301 TGGAGCAATG AAAAATGAT GCGGGAAGTT GCGCCCGATG CCTTCAGCGG
 351 CAATCCTGAA GGGCAGTTTT TTCCCAGACG CTACGAAATC GATGCGGGCG
 401 GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCCGC
 20 451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
 501 TTATGAAATG CTGATATATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
 551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAATC
 601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
 651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCCGCAC ACGCCGTACA
 25 701 aCAcTatac gggcgggggc ttgccgccaa cccgattgc gctgcccgC
 751 AaggcggaAa tggatgccgc cgcccaccgc tccggcgAa aatacctgTa
 801 ttctgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
 851 TGACCGAACA CAACGCCGCc gTcCGCAAAT ATATTTTGAA AAAATAA

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

30 1 ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
 51 PSEVSAWDIL QKMRGGRPDS VTVQIEGSR FSHMRKVIDA TPDIGHDTKG
 101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR
 151 LINEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI
 201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPYNTYTGGG LPPTRIALPG
 35 251 KAAMDAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNA VRKYILKK*

ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

	10	20	30	40	50	60
orf7-1.pep	KLLKWSAVFLTVSAAVF	FAALLFV	PKDNGRAYRIKIAKNQGI	SSVGRKLAEDRIVFSRHVL		
orf7ng-1				YRIKIAKNQGI	SSVGRKLAEDRIVFSRHVL	
				10	20	30
	70	80	90	100	110	120
orf7-1.pep	TAAAYVLGVHNRLHTGT	YRLPSEVSAWDILQKMRGGRPDSVT	VQIEGSRFSHMRKVIDA			
orf7ng-1	TAAAYVLGVHNRLHTGT	YRLPSEVSAWDILQKMRGGRPDSVT	VQIEGSRFSHMRKVIDA			
	40	50	60	70	80	90
	130	140	150	160	170	180
orf7-1.pep	TPDIGHDTKGWSNEKLMAEV	APDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR				
orf7ng-1	TPDIGHDTKGWSNEKLMAEV	APDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR				
	100	110	120	130	140	150
	190	200	210	220	230	240
orf7-1.pep	LNEAWESRQDGLPYKNPYE	MLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV				
orf7ng-1	LNEAWAGRQDGLPYKNPYE	MLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV				
	160	170	180	190	200	210
	250	260	270	280	290	300
orf7-1.pep	IYGMGAAYKGKIRKADLRD	TPYNTYTGGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS				

sp|P28306|YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
has 97 additional C-terminal residues [Escherichia coli] Length = 340
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 20/87 (22%), Positives = 40/87 (45%)

```
Query:      70 SVTVQIIIEGSRFESHMRKVIDATPDIGH 96
            ++++EG R S   K +   P I H
Sbjct:     109 OFPLRLVEGMRLSDYLKOLREAPYIKH 135
```

Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEK 179
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
Sbjct: 158 EGWFPDPTWMTYANTTDDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNOLVTMASIIIEK 217

Query: 180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRDTPYNTYTTGG 239
ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

```
Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYL YFVSKMDG 274
      GLPP IA PG ++ AAAHP+ YLYFV+ G
Sbjct: 278 GLPPGAIAITPGADSLKAAAHPAKTPYLYFVADGKG 312
```

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

1	CGTTTCAAAA	TGTTAACTGT	GTTGACGGCA	ACCTTGATTG	CCGGACAGGT
51	ATCTGCCGCC	GGAGGCGGTG	CGGGGATAT	GAACAGCCG	AAGGAAGTCG
101	GAAAGGTTTT	CAGAAAGCAG	CAGCGTTACA	GCGAGGAAGA	AATCAAAAAC
151	GAACGCGCAC	GGCTTGGCGC	AGTGGGTCAG	CGGGTTAATC	AGATATTTCT
201	TGTGCTGGGA	GGGGA AACCG	CTTGTCAAAA	GGGGCAGCGG	GGAAACGGCTC
251	TGGCAACCTA	TATGCTGATG	TTGGAACGCA	CAAAATCCCC	CGAAGTCGCC
301	GAACGCGCCT	TGGAATGGC	CGTTCGCTG	AACGCGTTTG	AACAGGCGGA
351	AATGATTAT	CAGAAATGGC	GGCAGATTGA	GCCTATACCG	GGTAAGGCGC
401	AAAAACGGCG	GGGTGGCTG	CGGAACCTGC	TGAGGGAAG	AGGA AATCAG
451	CATCTGGACG	GACGGGAAGA	AGTGTGGCT	CAGGCGGACG	AAGGACAG

This corresponds to the amino acid sequence <SEO ID 40; ORF9>:

```

1  ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFQRKQ QRYSEEEIKN
51 ERARLAAVGE RVNQIFTLIG GETALQKGOA GTALATYMLM LERTKSPEVA
101 ERALEMAVSL NAFEQAEMII QKWROIEPII GKAKRAGWL RNVLRERGNO

```

151 HLDGREEVLA QADEGQ

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

```

1  ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT
51  TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
101 CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA
151 GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA
201 TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGCAGG
251 CGGGAACGGC TCTGGCAACC TATATGCTGA GTTGGAAACG CACAAAATCC
301 CCCGAAGTCG CCGAACGCGC CTTGGAATG GCCGTGTCGC TGAACGCGTT
10  351 TGAACAGCGC GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
401 CGGGTAAGGC GCAAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA
451 AGAGGAAATC AGCATCTGGA CGGACTGGAA GAAGTGCTGG CTCAGGCGGA
501 CGAAGGACAG AACCGCAGGG TGTTTTATT GTTGGCACAA GCCGCCGTGC
551 AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTGC CCGCGCGGCG
15  601 TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTTCA
651 CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG
701 CGAAGCTCGA TACGGAAATA TTGCCCCCA CTTAATGAC GTTGCCTCTG
751 ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
801 CACCCAAAAC CTTTCGGCGG TCTGGCAGGA AATGGAAATT ATGAATCTGG
20  851 TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG
901 GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
951 AAACCGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
1001 ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG
1051 ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
25  1101 AAAAGTATCC GCGCCGGAAT ACCTGTTTGA CAAAGGTGTG CTGGCGGCTG
1151 CGGCGGCTGT CGAGTTGGAC GCGGCGAGG CGGCTTTGCG GCAGATCGGC
1201 AGGGTGCGGA AACTTCCCGA ACAGCAGGG CGGTATTTTA CGGCAGACAA
1251 TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
1301 AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAAACGCC TGCCGGCAGT
30  1351 AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTGAG TTGTTTACGA
1401 TCGGCTTGGC AAGCGGAAAA AAATGATTTC AGATCTTGAA AGGGCGTTCA
1451 GCGTTGCAAC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
1501 CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
1551 ATACCAAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
35  1601 CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
1651 TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCAT TGGGCGAAGT
1701 GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTGACGTA TGGACGCAGG
1751 CGGCACACCT TACGGGAGAC AAGAAAATAT GCGGGGAAAC GCTCAAACGT
1801 CACGGCATCG CATTGCCCA ACCTTCCCGA AAACCTCGGA AATAA

```

40 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

```

1  MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQORYSEE
51  EIKNERARLA AVGERVNQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE
45  151 RGNQHLGLE EVLAQADEGQ NRRVFLLLAQ AAVQQDGLAQ KASKAVRRRA
201 LKYEHLPEAA VADVVFVQVG REKEKAIGAL QRLAKLDTEI LPPTLMTLRL
251 TARKYPEILD GFFEQTDTQN LSAVWQEMEI MNLVSLHRLD DAYARLNVLL
301 ERNPNADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTA
351 MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAVELD GGRAALRQIG
401 RVRKLPEQQG RYFTADNL SK IQLALSKLP DKREALRGLD KIIEKPPAGS
50  451 NTELQAEALV QRSVVYDRIG KRKKMISDLE RAFRLAPDNA QIMNNLGYSL
501 LTDSKRLDEG FALLQTAYQI NPDDTAVNDS IGWAYYLGKD AESALPYLRY
551 SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR
601 HGIALPQPSR KPRK*

```

Computer analysis of this amino acid sequence gave the following results:

55 Homology with a predicted ORF from *N.meningitidis* (strain A)ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N.**meningitidis*:

```

60  orf9.pep      10      20      30      40      50
                RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQORYSEEEIKNERARLA
                || :||:|:|:|:|:| || ||:| | |||||:|:|:|:|:|:|:|:|:|:|:|:|:|
orf9a          MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQORYSEEEIKNERARLA

```


-80-

		10	20	30	40	50
	60	70	80	90	100	110
5	orf9.pep	AVGERVNIQIFTLGGETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
	orf9a	AVGERVNIQIFTLGKETALQKQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA				
		60	70	80	90	100
10	orf9.pep	120	130	140	150	160
	orf9a	EMIQKWRQIEPIPGKAQKRAQWLRNVLRRGNQHLDGREEVLAQADEGQ				
		120	130	140	150	160
15	orf9a	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQXREKEKAIGALQRLAKLDEI				
		180	190	200	210	220
						230

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

1	ATGTTACCCG	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
51	TGCCGGGCAG	GCGTATGCCG	CGGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
101	AAGTCGGAAG	GGTTTTCAGA	AAGCAGCAGC	GTTACAGCGA	GGAAGAAATC
151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCGAA
301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
351	GGCGGAAATG	ATTTATCAGA	AATGGCCGCA	GATTGAGCCT	ATACCGGGTA
401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAGAGGGA
451	AATCAGCATC	TAGACGGACT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
501	ACAGAACCGC	AGGGTGTTTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
551	ACGGGTTGGC	GCAAAAAGCA	TCGAAAGCGG	TTCGCCGCGC	GGCGTTGAGA
601	TATGAACATC	TGCCCGAAGC	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	CTGACTGCA
751	CGCAAAATATC	CCGAAATACT	CGACGGCTTT	TTGAGCAGCA	CAGACACCCA
801	AAACCTTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
851	TGCACAGGCT	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
901	AATCCGAATG	CAGACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAAAACCG
951	AAAAGAANGT	GCTTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
1051	TATGCCGACC	GAAGGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
1101	GTCGCGCGCG	GAATACCTGT	TCGACAAAAG	TGTGCTGGCG	GCTGCGGCGG
1151	CTGTGAGATT	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
1251	CAAAATACAG	ATGTTCCGCC	TGTCGAAGCT	GCCCGACAAA	CGGAGGCTT
1301	TGAGGGGGTT	GGACAAGATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTTACGGCTT
1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
1501	GATTCCAAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTTCAGA	CGGCATACCA
1551	AATCAACCCG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
1601	ACCTGAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCTGTT
1651	GAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTG
1701	GGCATTGGGC	GAACGCGATC	AGGCGGTGTA	CGTATGGACG	CAGGCGGCAC
1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTCACGGC
1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

1	MLPARFTILS	VLAALLLAGQ	AYAAGAADAK	PPKEVGKVFR	KQORYSEEEI
51	KNERARILAAV	GERVNIQIFTL	LGXETALQKG	QAGTALATYM	LMLERTKSPE
101	VAERALEMAV	SLNAFEQAEM	IYQKWRQIEP	IPGKAQKRAQ	WLRNVLRRER
151	NQHLDGLEEX	LAQADEXQNR	RVFLLLAQAA	VQQDGLAQKA	SKAVRRAALR
201	YEHLPEAAVA	DVVSFVQXRE	KEKAIGALQR	LAKLDEILP	PTLMTLRLTA
251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSLHRLDDA	YARLNVLLER
301	NPNADLYIQA	AILAANRKEK	ASVIDGYAEK	AYGRGTGEQR	GRAAMTAAMI
351	YADRRDYTEK	RQWLKKVSAP	EYLFDKGVLA	AAAARELDXG	RAALRQIGRV
401	RKLPEQQGRY	FTADNLSKIQ	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
451	ELQAEALVQR	SVYDRLGKR	KMISDLERA	FRLAPDNAQI	MNLLGYSLLS
501	DSKRLDEGFA	LLQTAYQINP	DDTAVNDSIG	WAYYLKXDAE	SALPYLRYSF
551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG

601 IALPQPSRKPK RK*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVF	10	20	30	40	50
	orf9-1	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPK	10	20	30	40	50
10	orf9a.pep	AVGERVNIQIFTLGKETALQKGAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	60	70	80	90	100
	orf9-1	AVGERVNIQIFTLGGETALQKGAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA	60	70	80	90	100
15	orf9a.pep	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHL	120	130	140	150	160
	orf9-1	EMIQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHL	120	130	140	150	160
20	orf9a.pep	AAVQDGLAQKASKAVRRAALRYEHLPEAAVADV	180	190	200	210	220
	orf9-1	AAVQDGLAQKASKAVRRAALRYEHLPEAAVADV	180	190	200	210	220
25	orf9a.pep	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL	240	250	260	270	280
	orf9-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL	240	250	260	270	280
30	orf9a.pep	ERNPNADLYIQAILAANRKEKXASVIDGYAEKAYGRGTGEQ	300	310	320	330	340
	orf9-1	ERNPNADLYIQAILAANRKEKASVIDGYAEKAYGRGTGEQ	300	310	320	330	340
35	orf9a.pep	KVRQWLKKSVAPEYLF	360	370	380	390	400
	orf9-1	KVRQWLKKSVAPEYLF	360	370	380	390	400
40	orf9a.pep	IQMFALSKLPDKREALRGLDKII	420	430	440	450	460
	orf9-1	IQMLALSKLPDKREALRGLDKII	420	430	440	450	460
45	orf9a.pep	RAFR LAPDNAQIMNNGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD	480	490	500	510	520
	orf9-1	RAFR LAPDNAQIMNNGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD	480	490	500	510	520
50	orf9a.pep	AESALPYLRYSFENDPEPEVA	540	550	560	570	580
	orf9-1	AESALPYLRYSFENDPEPEVA	540	550	560	570	580
55	orf9a.pep	HGIALPQPSRKPRKX	600	610			
	orf9-1	HGIALPQPSRKPRKX	600	610			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae*:

5	Orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVGVFRKQORYSEEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVELPKEVGKVLKRHRRYSEEEIKNERAR	58
10	orf9	LAAVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPVAERALEMAVSLNAFE	118
	orf9	QAEMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPASDYVHQPMIFLLL	178

- 15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

20	1	MIMLPARFTI	LSVLAAALLA	GQAYAAGAAD	VELPKEVGKV	LRKHRRYSEE
	51	EIKNERARLA	AVGERVNRVF	TLLGGETALQ	KGQAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIQKWRQI	EPIPGEAQKP	AGWLRNVLKE
	151	GGNPHLDRLE	EVPAQSDYVH	QPMIFLLLVQ	AAVQHGGVAQ	KPSKAVRPAA
	201	YNYEVLPEA	GADAVFCVQG	PQYEKAIQSF	PPCGRNPQTE	NIAPPFNELF
	251	RPTARPISPK	LLQRFRTTEP	NLAKPFRPPG	PEMETYQTGF	PRPLTRNNPT

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

- 25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

30	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCTCGCAG	CAGCCCTGCT
	51	TGCCCGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAA	GTTTAAAG	AAACATCGGC	GTTACAGCGA	GGAAGAAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAACGGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
35	301	GTCCGCCGAC	GCGCCTTGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATG	ATTTATCAGA	AATGgcggca	gatcgagcct	ataCcggtg
	401	agcgcaaaaa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGga
	451	aaTCAGCATC	TGGAacgggtt	gaaagaggTG	CtggcgcaAT	cggacgatGT
	501	GCAAAAACgc	aggaTATTTT	TGCTGCTGGT	GCAAGCCGCC	GTGCagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTCCGcgtgc	GGcgttgaAG
40	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCC	CCCACTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAATATC	CCGAAATACT	CGACGGCTTT	TTCCGAGCAGA	CAGACACCCA
	801	AAACCTTTTCG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	TGCGTAAGCC	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACAC
45	901	AACCCGAATG	CAAACCTGTA	TATTCAAGCG	GCGATATTGG	CGGCAAACCG
	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATgacggc	GGCGATGATA
	1051	TATGCCGACC	GCAGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	CGTGCTGGCG	GCTGCGGCGG
	1151	CTGCCGAATT	GGACGAGGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
50	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTGTGC
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	GCCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
	1351	GAACCTTTGG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
	1401	cggCAACCGG	GGA AAAATGA	TTGCCGACCT	tgaAACcgcg	CTCAAACCTTA
	1451	CGCCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
55	1501	GATTCCAAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAACGA	CAGCATAGGC	TGGGCGTATT
	1601	ACCTGAAAGG	CGACgcggaA	AGCGCGCTGC	CGTATCTGcg	gtattcgttt
	1651	gAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTGTG

-83-

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5 1 MLPARFTILS VLAAALLAGO AYAAGAADVE LPKEVGKVLK KHRRYSEEEI
 51 KNERARLAAV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE
 101 VAERALEMAV SINAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG
 151 NQHLDGLKEV LAQSDDVQKR RIFLLLVQAA VQGGGVAQKA SKAVRRAALK
 10 201 YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA
 251 RKYPEILDGF FEQTDTONLS AVWQEMEIMN LVSLRKPDDA YARLNVLEH
 301 NPNANLYIQA AILAANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
 351 YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV
 401 RKLPEQQGRY FTADNLSKIQ MLALSCLPKD REALIGLNNI IAKLSAAGST
 451 EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLTDPDPAQI MNNLGYSLLS
 15 501 DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLGDAE SALPYLRYSF
 551 ENDPEPEVAA HLGVLWALG ERDQAVDVT QAAHLRGDKK IWRETLKRYG
 601 IALPEPSRKP RK*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	10	20	30	40	50	60
		MLPNRFKMLTVLTATLIAGQVSAAGGAGDMKQPKVEGKVFVRKQORYSEEEIKNERARLA					
	orf9ng-1	MLPARFTILSVLAAALLAGQAYAG--AADVELPKEVGKVLKHHRRYSEEEIKNERARLA					
		10	20	30	40	50	
25	orf9-1.pep	70	80	90	100	110	120
		AVGERVNQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
	orf9ng-1	AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA					
		60	70	80	90	100	110
30	orf9-1.pep	130	140	150	160	170	180
		EMIIYQKWRQIEPIPGKAQKRAWLRNVLRERGNQHLGLEEVLAQADEGQNRVFLLLAQ					
	orf9ng-1	EMIIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLKEVLAQSDDVQKRIFFLLVQ					
35		120	130	140	150	160	170
40	orf9-1.pep	190	200	210	220	230	240
		AAVQQDGLAQKASKAVRRAALKYEHLEPAAVADVFSVQGREKEKAIGALQRLAKLDTEI					
	orf9ng-1	AAVQQGGVAQKASKAVRRAALKYEHLEPAAVADAVFGVQGREKEKAIEALQRLAKLDTEI					
		180	190	200	210	220	230
45	orf9-1.pep	250	260	270	280	290	300
		LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL					
	orf9ng-1	LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKPDDAYARLNVLL					
		240	250	260	270	280	290
50	orf9-1.pep	310	320	330	340	350	360
		ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQRRAALTAAMMYADRRDYA					
	orf9ng-1	EHNPNANLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYA					
		300	310	320	330	340	350
55	orf9-1.pep	370	380	390	400	410	420
		KVRQWLKKVSAPEYLFDKGVLA AAAA VELDGGRAALRQIGRVRLKPEQQGRYFTADNLSK					
	orf9ng-1	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVRLKPEQQGRYFTADNLSK					
		360	370	380	390	400	410
60	orf9-1.pep	430	440	450	460	470	480
		IQMLALSCLPKDREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRRKMISDLE					
	orf9ng-1	IQMLALSCLPKDREALIGLNNIIAKLSAAGSTEPLAEALAQRSSIYEQFGKRGKMIADLE					
65		420	430	440	450	460	470
		490	500	510	520	530	540

-84-

```

orf9-1.pep  RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLGKD
              |::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf9ng-1     TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLGKD
              480      490      500      510      520      530

5           550      560      570      580      590      600
orf9-1.pep  AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTOAAHLTGDKKIWR ETLKR
              |::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf9ng-1     AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTOAAHLRGDKKIWR ETLKR
              540      550      560      570      580      590

10          610
orf9-1.pep  HGIALPQPSRKPRKX
              :||||:|||||
orf9ng-1     YGIALPEPSRKPRKX
              600      610

```

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```

sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
(ORF3)
>gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
(X82071) orf3 [Pseudomonas aeruginosa] Length = 576
Score = 128 bits (318), Expect = 1e-28
Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

```

```

Query: 67  VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIQKWR 126
          +++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
Sbjct: 53  LYSLLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSEAFRAIEYLGADQEALDTSLLWA 112

```

```

Query: 127 QIEPIPGEAQKPAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
          + P +AQ+ A ++ VL G+ H D L A++D + +
Sbjct: 113 RSAPDNLDQAQAAAIQLARAGRYEESMVYMEKVLNGQGDT HFDLALSA AETDPDTRAGL 172

```

```

Query: 173 FXXXXXXXXXXXXXXXXKASKAVRRAALKYEHLP EAAVADAVFGVQGREKEKAIEALQRLA 232
          ++ KY + + A+ Q ++A+ L+ +
Sbjct: 173 L-----QSFHLLKKYPNNQQLLFGKALLLQODGRPDEALTLEDNS 214

```

```

Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLRKP 287
          E+ P L + L + K P + G E D + + + + LV +
Sbjct: 215 ASRHEVAPLLLRSLQLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

```

```

Query: 288 DDAYARLNVLL EHNPN-----ANLYIQAAI----- 312
          DDA A L++ P+ A +Y++ +
Sbjct: 271 DDAKA EFAGLVQQFPDDDDDLRFSLALVCLEAQAWDEARIYLEELVERD SHVDAAHFNLG 330

```

```

Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQGRGAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
          LA +K+ A +D YA+ G G + T ++ A R D A R + P+
Sbjct: 331 RLAEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

```

```

Query: 372 YLFDKXXXXXXXXXXXXXXXXXQRIGRVRKLP EQQGRYFTADNLSKIQLALSKLPDKR 431
          Y A L I+ ALS +
Sbjct: 389 Y-----AIQLYLIEAEALSNNDDQOE 408

```

```

Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
          +A + + + E L L RS++ E+ +M DL + PDNA +
Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTRSM LAEKRNDLAQMEKDLRFVIAREPDNAMAL 462

```

```

Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLGKDAESALPYLRY SFE 551
          N LGY+L + R E L+ A++NPDD A+ DS+GW Y +G A YLR + +
Sbjct: 463 NALGYTLADRTRYGEARELILKAHKLNPDPAILD SMGWINYRQGLADAERYLRQALQ 522

```

```

Query: 552 NDPEPEVA AHLGEVLWALGERDQAVDVWTOAAHLRGDKKIWR ETLKR 598
          P+ EVA AHLGEVLWA G + A +W + + D + R T+KR
Sbjct: 523 RYPDHEVA AHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

```

```

gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
Score = 81.5 bits (198), Expect = 1e-14
Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

```

```

Query: 408 GRYFTADNL-SKIQLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ----- 459
          G Y A L K ++LA PDK+E L + +K + + L +

```

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
 +I+Y+ G L A++L P+N N LGYSL +R++E L++

5 Sbjct: 391 VYFMEAIIVYDNLGDIKNAEKALRKAIELDPENPDYNNYLGYSLLLWYGKERVEEAEELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYLKGAESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572
 A + +P++ A DS+GW YYLKG D E A+ YL + E +P V H+G+VL +G +

10 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKG DYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510

Query: 573 DQAVDVWTQAAHLRGDKK 590
 ++A + + +A L + K

Sbjct: 511 EARNYYERALKLLEEGK 528

- 15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
 151 TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
 201 GACCAACGCC TCTTACCCTT CTATGGCGAA AATGCGTGCC GCCGCACCCA
 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
 25 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CG_aCTGGGCG
 351 GCTGCCGTGCC TATGCTGTG CAAATCCCCG TCTTCATCGG ATGTATTGG
 401 GCATTGTTTC CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT
 451 TACCGACCTC AGCCGCGCGC ACCCCTACTA CATCCTGCCC ATCATTATGG
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCGCGCGCC GACCGACCG
 30 551 ATGCAGGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGT
 601 CTTCTTCTTC CCTGCCGgks TGGTATTGTA CTGGGTAGTC AACAACTCC
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 ..NLYAGPQTS VIANIADNLQ LAKDYGKVHW FASPLEWLLN QLHNIIGNWG
 51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQW
 101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LROAPWLGWI
 151 TDLSRADPYI ILPII MAATM FAQTYLNPPP TDFMQAKMMK IMPLVFSXXF
 201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

- 40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GCGTTTTTTC GCCATCGCGC TGGTGATTAT
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
 101 AACAGGCAGC ACAACAACAG GCGTAACCG CTCCGCGCA AGCCGCGCTC
 151 GCGCCCGCAA CGCGGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
 45 201 TGATGAAAAA AGCGGCGACC TCGCGCGGCT GACCCTGCTC AAATACAAAG
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGCGCA CGGCAAAGAA
 301 TACACCTAGC TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
 50 451 ATCGACAAAG TTTATACTTT CACCAAGGC AGCTATCTGG TCAACGTCCG
 501 CTTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551 ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTACTTTACC
 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
 701 CCGAATACAT CCGAACCAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
 751 CACTTCATGT CCACCTGGAT TCTCAACCT AAAGGCAGAC AAAGCGTTTG
 801 CGCGCGAGGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
 851 ACAGCACCAG CGTCAGCGTG CCTTTAGCGC CCATCCAAAA CGGCGCGAAA
 901 CGCGAAGCCT CCATCAACCT CTACGCGCGC CCGCAGACCA CATCCGTCAT
 60 951 CGCAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160

1001 ACTGGTTCGC CTCCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCGCCGC ACCCAAATG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAAACAGC GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GCGCGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTCGC CCAAATATAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAATCATAT GCCGTGGGTT
1501 TTCTCCGTCA TGTTCCTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAAACA ACGCGCCCAA GCGAAGTCG TTTCTTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15
20
25
30
35
40
45
50

1 MDFKRLTAFF AIALVIMIGW EKMFPPTPKPV PAPOQAAQQQ AVTASAEAL
51 APATPTITVT DTQVAVIDEK SGDRLRLTLL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAQGNILKLG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVYTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVYVT PEGNFQKVSF SLDLDDAKSG KSEAEYIRKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHWFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEYGDD
401 RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPPYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 FSVMFFFFPA GLVLYWVVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600

ORF11 2 LYAGPQTTSVIANIADNLQAKDYGVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIIVK 61
LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
60K 324 LYAGPKIQSKLSPGLELTVDYGFLLWFTAQPIFWLLQHIHSLGNWGWSIIVLTMLIK 383
ORF11 62 AVLYPLTNASYRSMAMRAAAPKLQAIKEYGDDRRXXXXXXXXXXLYTDEKINPLGGCLPM 121
+ +PL+ ASYRSM+MRA APKL A+KE++GDDR LY EKINPLGGCLP+
35 60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443
ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRDPYILPIIIMATMFAQTYLNPPPT 181
L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
40 60K 444 LVQMPVFLALYWVLLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQRLNPTTP 503
ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXPAQXVLYWVVNNLLTIAQQWHINRSIE 230
DPMQAK+MK+MP++ PAG VLYWVVNN L+I+QQW+I R IE
45 60K 504 DPMQAKVMKMPIIFTFFFLWFPAGLVLYWVVNNCLISQQWYITRRIE 552

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N.*

meningitidis:

50
55

orf11.pep
orf11a
orf11.pep
orf11a

280 290 300 310 320 330
340 350 360 370 380 390

10 20 30
40 50 60 70 80 90

NLYAGPQTTSVIANIADNLQAKDYGVHWF
IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQAKDYGVHWF
FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE
FASPLFWLLNQLHNIIGNWGWAIIVLTIIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE

-87-

		100	110	120	130	140	150
orf11.pep		KYGD	DRMAQQQ	AMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI			
orf11a		KYGD	DRMAQQQ	AMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI			
5		400	410	420	430	440	450
		160	170	180	190	200	210
orf11.pep		TDLSRADPYYILPI	IMAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXFFFPAGXVLY				
10		TDLSRADPYYILPI	IMAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXFFFPAGXVLY				
		460	470	480	490	500	510
		220	230	240			
orf11.pep		WVNNLLTIAQQWHINRSIEKQRAQGEVVSX					
15		WVNNLLTIAQQWHINRSIEKQRAQGEVVSX					
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

	1	ANGGATTTTA	AAAGACTCAC	NGNGTTTTTC	GCCATCGCAC	TGGTGATTAT
20	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCCGCGCCCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCCGCGCA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGGCT	GACCCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CNAAAATAAA	CCGTTTCATCC	TGTTTGCGCA	CGGCAANAA
25	301	TACACCTACN	TCGCCCANTC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTCGG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
30	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCGG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCGACTTGG	ACGACGATGC	CAANTCCGGN	AAATCCGAGG
	701	CGGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
	751	CACTTCATGT	CCACCTGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
35	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACACAG	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCCNAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAAAATC	GCCGACAACC	TGCAACTGGN	CAAAGACTAC	GGCAAAGTAC
40	1001	ACTGGTTTCG	CTCCCCCTC	TTTTGGCTTT	TGAACCAACT	GCACAACATC
	1051	ATCGGCAACT	GGGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCCTCTTA	CCGTTTCGATG	GCGAAATGTC
	1151	GTGCCGCCGC	GCCCAAACTG	CAAGCCATCA	AAGAGAAATA	CGGCGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAT
	1251	CAACCCGCTG	GGCGGCTGCC	TGCCTATGCT	GTTGCAATC	CCCGTCTTCA
45	1301	TCGGATTGTA	TTGGGCATTG	TTGCCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTTCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
	1501	NTNTCNNNNA	NGTTCCTCNN	CTCCCTGCC	GGTCTGGTAT	TGTACTGGGT
50	1551	GATCAACAAC	CTCCTGACCA	TCGCCAGCA	ATGGCACATC	AACCGCAGCA
	1601	TCGAAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCCTAA	

This encodes a protein having amino acid sequence <SEQ ID 54>:

	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPQQTAAQQ	AVXASAEAL
55	51	APXXPITVTT	DTVQAVIDEK	SGDLRLRLTL	KYKATGDXNK	PFILFGDGKX
	101	YTYXAXSELL	DAQGNNILKG	IGFSAPKKQY	SLEGDKVEVR	LSAPETRGLK
	151	IDKVYTFTKG	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSVVGPPVYT	PEGNFQKVSF	SDLDDAXSG	KSEAERYIRKT	XTGWLGMIEH
	251	HFMSTWILQP	KGGQSVCAAG	DCXXDIKRRN	DKLYSTSVSV	PLAAIQNGAK
	301	SXASINLYAG	PQTTSVIANI	ADNLQLXKDY	GKVHWFASPL	FWLLNQLHNI
60	351	IGNWGWAIIV	LTIIVKAVLY	PLTNASYRSM	AKMRAAAPKL	QAIKEYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLRS	ADPYYILPII	MAATMFAQTY	LNPPPTDPMQ	AKMMKIMPLV
	501	XSXXFFXFPA	GLVLYWVINN	LLTIAQQWHI	NRSIEKQRAQ	GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

65	10	20	30	40	50	60
----	----	----	----	----	----	----

-88-

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXXMFPTPKPVPAPQQTAAQQAVXASAEALAPXXPITVTT
	orf11-1	MDFKRLTAAFFAIALVIMIGWEKMFPTPKPVPAPQQAQQAAVTASAEALAPATPITVTT 10 20 30 40 50 60
10	orf11a.pep	70 80 90 100 110 120 DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFFILFGDGKXYTYXAXSELLDAQNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDNKPFILFGDGKEYTYVAQSELLDAQNNILKG 70 80 90 100 110 120
15	orf11a.pep	130 140 150 160 170 180 IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYFTFKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYFTFKGSYLVNVRFDIANGSGQTANL 130 140 150 160 170 180
20	orf11a.pep	190 200 210 220 230 240 SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAERYIKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAERYIKT 190 200 210 220 230 240
25	orf11a.pep	250 260 270 280 290 300 XTGWLGMIEHHFMSTWILQPKGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK 250 260 270 280 290 300
30	orf11a.pep	310 320 330 340 350 360 SXASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV :
	orf11-1	AEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV 310 320 330 340 350 360
35	orf11a.pep	370 380 390 400 410 420 LTIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
	orf11-1	LTIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL 370 380 390 400 410 420
45	orf11a.pep	430 440 450 460 470 480 GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAATMFAQTY 430 440 450 460 470 480
50	orf11a.pep	490 500 510 520 530 540 LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ 490 500 510 520 530 540
55	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

60 Homology with a predicted ORF from *N.gonorrhoeae*ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N.**gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT 57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT 60

	orf11	IIVKAVLYPLTNASYRSMKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMKMRAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN	180
10	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGE	240
	orf11	VVS	240
15	orf11ng	VVS	243

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

	1	MAVNLYAGPQ	TTSVIANIAD	NLQAKDYGK	VHWFASPLFW	LLNQLHNIIG
20	51	NWGWAIVVLT	IIVKAVLYPL	TNASYRSMK	MRAAAPELQT	IKEKYGDDRM
	101	AQQQAMMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
	201	VMEFFFPAGL	VLYWVVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

25	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCCGA	AGCCGCGCTC
	151	CGCCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTAT
	201	TGATGAAAAA	AGTGGCGACC	TGCGCCGGCT	GACCTTGCTC	AAATACAAAG
30	251	CAACCGCGCA	CGAAAAACAA	CCGTTTCGTC	TGTTTGCGCA	CGGCAAGAA
	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	ACCCTCAACG
	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGC	CGGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAGAC	AGCTATCTGG	TCAACGTCGG
35	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
	551	ACCGCATCGT	CCGCGACCAC	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	ccgaatacat	CCGCAAAACC	ccgaccggtt	ggctcggcat	gattgaacac
40	751	cacttcatgt	ccacctggat	cctccAAcct	aaaggcggcc	aaaacgtttg
	801	cqccccagga	gactgccgta	tcgacattaa	aCgccgcaac	gacaagctgt
	851	acagcgcaag	cgtcagcggt	cctttaaccg	ctatcccaac	ccggggggcca
	901	aaaccgaaaa	tggcggTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
	951	TATCGCAAAC	ATCGCcgacA	ACCTGCAACT	GGCAAAAGAC	TACGGTAAAG
45	1001	TACACTGGTT	CGCATCGCCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcctc	ctACCGTTCG	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgGCGAC
	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAgacgAGAA
	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctggttCAA	ATCCCCGTCT
50	1301	TCATCGGCTT	GTACTGGGCA	TTGTTTCGCCT	CCGTAGAATT	GCGCCAGGCA
	1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT
	1401	CCTGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAAACC	TATCTGAACC
	1451	CGCCGCCGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAAT	CATGCCGTG
	1501	GTTTTCTCCG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTCTCTACTG
55	1551	GGTGGTCAAC	AACCTCCTGA	CCATCGCCCA	GCAGTGGCAC	ATCAACCGCA
	1601	GCATCGAAAA	ACAACGCGCC	CAAGGCGAAG	TCGTTTCCTA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

60	1	MDFKRLTAFF	AIALVIMIGW	EKMFPPTPKPV	PAPQQAQKQ	AATASAEAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRLRLTL	KYKATGDENK	PFVLFEGDKE
	101	YTYVAQSELL	DAQGNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYTFETD	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQAKD	YGKVHWFASP	LFWLLNQLHN

-90-

351 IIGNWGWAI VLTIIIVKAVL YPLTNASYRS MAKMRRAAPK LQTIKEYGD
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
 451 PWLGWITDLS RADPYIILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
 501 VFSVMFFFFP AGLVLYWVNV NLLTIAQQWH INRSIEKQRA QGEVVS*

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAAQQAATASAEALAPATPITVTT					
10	orf11-1	MDFKRLTAFFAIALVIMIGWEKMFPTPKVPVPAQQAAQQAATASAEALAPATPITVTT					
		10	20	30	40	50	60
	orf11ng-1.pep	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
15	orf11-1	DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
		70	80	90	100	110	120
	orf11ng-1.pep	IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYFTTKDSYLVNVRFDIANGSGQTANL					
20	orf11-1	IGFSAPKKQYSLGDKVEVRLSAPETRGLKIDKVYFTTKDSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLDDAKSGKSEAERYKRT					
25	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVYTPEGNFQKVSFSDLDDAKSGKSEAERYKRT					
		190	200	210	220	230	240
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDRCRIDIKRRNDKLYSASVSVPLTAIPTRGP					
30	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA					
		250	260	270	280	290	
	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAI					
40	orf11-1	KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAI					
		300	310	320	330	340	350
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP					
45	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP					
		360	370	380	390	400	410
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMATMFAQT					
50	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMATMFAQT					
		420	430	440	450	460	470
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVNVNLLTIAQQWHINRSIEKQRA					
55	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVNVNLLTIAQQWHINRSIEKQRA					
		480	490	500	510	520	530
	orf11ng-1.pep	QGEVVSX					
60	orf11-1	QGEVVSX					
		540					

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

-91-

ID 60IM_PSEPU STANDARD; PRT; 560 AA.
 AC P25754;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES Init1: 1074 Initn: 1293 Opt: 1103
 Smith-Waterman score: 1406; 41.5% identity in 574 aa overlap

		10	20	30	40
orfl1ng-1.pep	MDFKR---LTAFFAIALVIMIGW-----EKMFT-----PKPVPAPQQAQKQ				
p25754	MDIKRTILIAALAVSVYVMVLKWNDDYGQAALPTQNTAASTVAPGLPDGVPAGNNGASAD				
		10	20	30	40
		50	60	70	80
orfl1ng-1.pep	AATASAEALAPATPIT-----VTDTVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF				
p25754	VPSANAESSPAELAPVALSKDLIRVKTQVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF				
		70	80	90	100
		110	120		
		100	110	120	130
orfl1ng-1.pep	VLFGDGKEYTYVAQSELLDAQGNILKIG---FSAPKKQYTL-NGD---TVEVRLSAPE				
p25754	QLFDNGGERVYLAQSGLTGTDGPDARASGRPLYAAEQKSYQLADGQELVVDLKFS---				
		130	140	150	160
		170			
		150	160	170	180
orfl1ng-1.pep	TNGLKIDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY				
p25754	DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKRDASGDPSSSTATGTATY				
		180	190	200	210
		220	230		
		210	220	230	240
orfl1ng-1.pep	VGPVVYTPPEGNFQKVSFSDLDLDDAKSGKSEAERYIKTPTGWLGMIEHHMSTWILQPKGG				
p25754	LGAALWTASEPYKKVSMKDID---KGSLE-----NVSGGWVAWLQHYFVTAWI-PAKSD				
		240	250	260	270
		280			
		270	280	290	300
orfl1ng-1.pep	QNVCAQGDCCRIDIKRRNDKLYSASVSVPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD				
p25754	NNV-----VQTRKDSQGYIIGYTGPIVSVPA-GGKVETSALLYAGPKIQSKLKELSP				
		290	300	310	320
		330	340	350	360
orfl1ng-1.pep	NLQLAKDYGVHWF-ASPLEWLLNQLHNIIGNWGWAIVVLTIIIVKAVLYPLTNASYRMA				
p25754	GLELTVDYGF-LWFIAQPIFWLLQHIHSLGNWGSIIIVLTMLIKGLFFPLSAASYRMA				
		340	350	360	370
		380	390		
		390	400	410	420
orfl1ng-1.pep	KMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF				
p25754	RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWLL				
		400	410	420	430
		440	450		
		450	460	470	480
orfl1ng-1.pep	ASVELRQAPWLGWITDLSRADPYIILPIIIMATMFAQTYLNPPTDPMQAKMMKIMPLVF				
p25754	ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQLNPTPPDPMQAKVMKMMPIIF				
		460	470	480	490
		500	510		
		510	520	530	540
orfl1ng-1.pep	SVMFFFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGEVVSX				
p25754	TFFFLWFPAGLVLYWVVNNCLSSISQWYITRRIEATKKA				
		520	530	540	550
		560			

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTtNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTtGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATtG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDL D AGQYVEILRH TGGNRYEVXY RGTxWQAQNT
101 GQEELEPGTR ALIVRKEGNI LIITHP*

```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20  . 1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GnCGCTCTGC TTTCCGCGCT GGGTATTtNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTtGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
25  251 ACCGTTACGA AGTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATtG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30  1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDL D AGQYVEILRH TGGNRYEVFY RGT#WQAQNT
101 GQEELEPGTR ALIVRKEGNI LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

35 *meningitidis*:

```

                                     10      20      30      40      50
orf13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
                |||||||||||||||||||||||||||||||||||||||
40  orf13a      MTVWFVA VAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
                10      20      30      40      50      60

                                     60      70      80      90     100     110
orf13.pep      VHAKTAVRK VETDSYQDL DAGQYVEILRH TGGNRYEVXYRGTxWQAQNTGQEELEPGTRA
                ||||||| ||||||| ||||||| ||||||| ||| ||||||| |||||||
45  orf13a      VHAKTAVGKVETDSYQDL DAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                70      80      90      100     110     120

                                     120
orf13.pep      LIVRKEGNLLIITHPX
                ||||||| |||
50

```

5
10

1	ATGACTGTAT	GGTTTGTTCG	CGCTGTTGCC	GTCTTAATCA	TCGAATTATT
51	GACGGGAACG	GTTTATCTTT	TGGTTGTCAG	CGCGGCTTTG	GCGGGTTTCGG
101	GCATTGCTTA	CGGGCTGACC	GGCAGCACGC	CTGCGCCCGT	CTTGACCCGC
151	GCTTGCCTT	CCGCGCTGGG	TATTTGGTTC	GTACACGCCA	AAACGCCGCT
201	GGCAAAAGTT	GAAACGGATT	CATATCAGGA	TTTGGATGCC	GGGCAATTATG
251	CCGAAATCCT	CCGGCACGCA	GGCGGCAACC	GTTACGAAGT	TTTTTATCGC
301	GGTACGCACT	GGCAGGCTCA	AAATACGGGG	CAAGAAGAGC	TTGAACCAGG
351	AACGCGCGCC	CTAATCGTCC	GCAAGGAAGG	CAACCTTCTT	ATCATCGCAA
401	AACCTTAA				

15 1 MTVWFVAAVA VLIIELTGT VYLLVSAAL AGSGIAYGLT GSTPAAVLTA
51 ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
101 GTHWOAONTG OEELEPGTRA LIVRKEGNLL IIAKP*

		10	20	30	40	50	60
20	orf13a.pep	MTVWFVA	AAVAVLII	IELLTGT	VYLLVVS	AALAGSG	IAYGLTG
	orf13-1		AVLIIEL	LTGTVYL	LVVSAAL	AGSGIAY	GLTGSTP
			10	20	30	40	50
25	orf13a.pep	70	80	90	100	110	120
		VHAKTAV	GVKETDS	YQDLDA	GQYAEIL	RHAGGN	RRYEVFY
	orf13-1	VHAKTAV	RKVETDS	YQDLDA	GQYVEIL	RHTGGN	RRYEVFY
		60	70	80	90	100	110
30	orf13a.pep	130	LIVRKEGNLLIIAKPX				
	orf13-1	LIVRKEGNLLIITHPX					
		120					

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N. gonorrhoeae*:

40	orf13	AVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTXXALLSALGIXF	51
	orf13ng	MTVWFVAAVAVLIIELLTGTVYLLVVSAAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF	60
45	orf13	VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVXRGTXWQAQNTGQEELEPGTRA	111
	orf13ng	VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFGRTHWQAQNTGQEVFEPGTRA	120
	orf13	LIVRKEGNLLIITHP	126
	orf13ng	LIVRKEGNLLLIANP	135

55

1	ATGACTGTAT	GGTTTGTTGC	CGCTGTTGCC	GTCTTAATCA	TCGAATTATT
51	GACGGGAACG	GTTTATCTTT	TGGTTGTCAG	CGCGGCTTTG	GCGGGTTCGG
101	GCATTGCCTA	CGGGCTGACT	GGCAGCACGC	CTGCCGCCGT	CTTGACCGCC
151	GCACTGCTTT	CGGCGCTGGG	CATTTGGTTT	GTACATGCCA	AAACGCCCGT
201	GGGAAAAGTT	GAAACGGATT	CATATCAGGA	TTTGATACCC	GGAAAATATG
251	CCGAAATCCT	CCGATACACA	GGCGGCAACC	GTTACGAAGT	TTTTTATCGC
301	GGTACGCACT	GGCAGGCGCA	AAATACGGGG	CAGGAAGTGT	TTGAACCGGG
351	AACGCGCGCC	CTCATCGTCC	GCAAAGAAGG	TAACCTTCTT	ATCATCGCAA
401	ACCCTTAA				

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10  orf13-1.pep      10      20      30      40      50
      AVLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTXALLSALGIXF
      |||||
orf13ng  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTAALLSALGIWF
      10      20      30      40      50      60

      60      70      80      90      100     110
orf13-1.pep VHAKTAVRKVETDSYQDL DAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
      |||||
15  orf13ng      VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
      70      80      90      100     110     120

      120
orf13-1.pep  LIVRKEGNLLIITHPX
      |||||
20  orf13ng      LIVRKEGNLLIIANPX
      130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that

25 ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30  1  ATGTwTGATT TCGGTTTrGG CGArCTGGTT TTTGTcGGCA TTATCGCCCT
      51  GATwGtCCTC GGCCcCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
      101 GGcTCATCGG CAGGCTGCAA CGCTTTGTcG GcAGCGTCAA ACAGGAATTT
      151 GACACTCAAA TCgAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
      201 AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
35  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
      301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
      351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
      401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40  1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
      51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
      101 LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45  1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTcGGCA TTATCGCCCT
      51  GATTGTCTC GGCCcCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
      101 GGTCATCGG CAGGCTGCAA CGCTTTGTcG GCAGCGTCAA ACAGGAATTT
      151 GACACTCAAA TCgAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
      201 AGTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
50  251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
      301 CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
      351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
      401 TGCCGTCCGA ACgTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
      451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
      501 GCGGGAATAC CTGACTGCTT CTGCCGCCCG ACCCGTCGTA CAGACCGTCG

```

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTC GTCCGAAACA
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5 1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNLP DAANTLSDGI SDVMPSEERSY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
 201 SLRKQAI SRK RDRPKHRAK PKLRVRKS*

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCTG GCAGCGTCAA ACAGGAATTT
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTGTA
 15 201 AGCTGCCGCT GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCCGAAC AGCGCACGCC TGCTGATTTC GGTGTGATG AAAACGGCAA
 351 TCCCTTTCCC GATGCGGCAA ACACCTTATT AGACGGCATT TCCGACGTTA
 401 TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCTTGG GGACAGCGGG
 20 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
 501 GCGGGAATAC CTGACTGCTT CTGCGCGCGC ACCCGTCGTA CAGACCGTCG
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTC GCATACCACT
 601 TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTC GTCCATAATC
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNFP DAANTLLDGI SDVMPSEERSY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
 30 201 SLRKQAI SRK RDLRPSRAK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELVFVGIIALIVL	GPERXPEAARXAGRLIGRLQRFVGSVKQEF	DTQIELEELR			
	orf2a	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR			
		10	20	30	40	50	60
40	orf2.pep	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS				
	orf2a	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP				
		70	80	90	100	110	120
45	orf2.pep	RCGKHPIRRHFRRYAV					
	orf2a	DAANTLLDGISDVMPSEERSY	ASAETLGDSGQTGSTAEP	ETDQDRAWREYLTASAAAPVV			
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR	60
	orf2-1	MFDFGLGELVFVGIIALIVL	GPERLPEAARTAGRLIGRLQRFVGSVKQEF	DTQIELEELR	60
55	orf2a.pep	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP		120
	orf2-1	KAKQEFEEAAAQVRDSLKET	GTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNLP		120
60	orf2a.pep	DAANTLLDGISDVMPSEERSY	ASAETLGDSGQTGSTAEP	ETDQDRAWREYLTASAAAPVV	180

orf2-1 DAANTLSDGI SDVMPERSYASAETLGDSGGTGSTAETDQDRAWREYLTASAAAPVV 180
 orf2a.pep QTVEVSYIDTAVETVPVHTTSLRKQAI SRKRDLPKSRAPKLRVRKXS 229
 5 orf2-1 QTVEVSYIDTAVETVPVHTTSLRKQAI SRKRDLPKSRAPKLRVRKXS 229

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 10 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
 101 LPEQRTPADF GVDEKGNLSL RYGKHRIRRH FRRYAV*

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT
 15 51 GATTGTCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
 101 GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
 151 GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCTGA
 201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA
 251 TGCGAAGCAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCGAAC AGCGCACGCC tgccgatttc gGTGTCGATg AAAacggcaa
 351 tcccccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
 401 TGCCGTCTGA ACCTTCCGAT ACTtccgcCG AAACCCCTTG GGACGACAGG
 451 CAAACCGGCA GTACAGCCGA ACCTGCCGAA ACCGACAAAG ACCGCGCATG
 501 GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg
 551 tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc
 601 acttccctgc gcaAACAGGC AATAAACCGC AAACCGGATT TttgtccgaA
 651 ACACCGCGCc aACCGAAat tgcgcgtcCG TAAATCATAA

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

1 MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 30 51 DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDQMNSLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGNPLP DTANTVSDGI SDVMPERSD TSAETLGDDR
 151 QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETVPVHT
 201 TSLRKQAINR KRDFCPKHRA KPKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

35 orf2.pep MXDFGLGELVFVGIIALIVL GPERXPEARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
 orf2ng MFDFGLGELIFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 40 orf2.pep KAKQFEAAAAAQVRDSLKETGDTMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS 120
 orf2ng KVKQAFEAAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNLSL 120
 orf2.pep RCGKHPIRRHFRRYAV 136
 45 orf2ng RYGKHRIRRHFRYAV 136

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

50 orf2-1.pep 10 20 30 40 50 60
 MFDFGLGELVFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
 orf2ng-1 MFDFGLGELIFVGIIALIVL GPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
 10 20 30 40 50 60
 55 orf2-1.pep 70 80 90 100 110 120
 KAKQFEAAAAAQVRDSLKETGDTMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
 orf2ng-1 KVKQAFEAAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP

-97-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf2-1.pep	DAANTLS	SDGISDV	MPSPERS	YASAETL	GDGSGQT	GSTAEP
	orf2ng-1	DTANTV	SDGISDV	MPSPERS	SDTSAETL	GDGDRQT	GSTAEP
		130	140	150	160	170	180
10	orf2-1.pep	Q-TVEVS	YIDTAV	ETPVPHT	TSLRKQ	AI	SRKRDF
	orf2ng-1	QRAVEVS	YIDTAV	ETPVPHT	TSLRKQ	AINRKR	DFCPKH
		190	200	210	220	230	

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),
 and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein
 of *E.coli*:

gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
 Score = 56.6 bits (134), Expect = 1e-07
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

Query: 1 MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +
 Sbjct: 1 MFDIGFSELLVFIIGLVVLGPRLPVAVKTVAGWIRALRSLATTQNELTQELKLQEFQ 60

Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
 +K+ +A+ + LK + +++ +
 Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane
 proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A
 shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results
 of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,
 whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis
 (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is
 a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

```

1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
51  CGC.TGCGGG AACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGTACTCC
251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
351 GTTTGACAGG TTTAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC
401 TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

```

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTCCTGCGC
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAANK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXXG EYINSPAVRT
 101 DYTYPYRSETT AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVP ANADTDVFIN
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG ACACGTACAG GTATTCCATC GCATGCGGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCGCTT CTGCCAGAGC TGCCGTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 15 201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCTGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 20 451 ATTGCGGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
 551 GCATAGACGT TGTCTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAAC TGAATATTTC GCAGTAGACA
 25 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
 801 AGGAATTAAG CCGACGGAAG GATTAAATGGT CGATTCTCC GATATCCGAC
 851 CATACGGCAA TCATACGGGT AACTCCGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA
 30 951 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAANK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPYRSETT AETTSGLTGT LTSLSTLNA PALSRQSDG SGSKSSLGLN
 35 151 IGGMGDYRNE TLTNPRDTA FLSHLVQTVF FLRGIDVVP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRNTKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSKEIK PTEGLMVDFA DIRPYGNHTG NSAPSVADN
 301 SHEGYGYSDE VVRQHRQGP *

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG ACACGTACAG GTATTCCATC GCATGCGGGA GGTAACGCT
 101 TTGCGGTCGA ACAAGAACTT GTGGCGCTT CTGCCAGAGC TGCCGTAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 45 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCTGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGCGGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 50 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
 551 GCATAGACGT TGTCTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAAC TGAATATTTC GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 55 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
 801 AGGAATTAAG CCGACAGAAG GATTAAATGGT CGATTCTCC GATATCCAAC
 851 CATACGGCAA TCATATGGGT AACTCTGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA
 951 AGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAANK

51 DMDLQALHGR KVALYIATMG DQSGSGLTGG RYSIDALIRG EYINSPAVRT
 101 DYTPRYETT AETTSGLTG LTSLSTLNA PALSRTQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVP ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKNAFEA
 251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
 301 SHEGYGSDE AVRRHRQGP *

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKMDLQALHGR
	orf15a	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKMDLQALHGR
15	orf15.pep	70 80 90 100 110 120	KVALYIATMGDQSGSGLTGGRYSIDAXXGEYINSPAVRTDYTPRYETTAETTSGLTG
	orf15a	70 80 90 100 110 120	KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
20	orf15.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15a	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
25	orf15.pep	190 200 210	FLRGIDVVPANADTDVFINIDVFGTIRNRTEM
	orf15a	190 200 210 220 230 240	FLRGIDVVPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKMDLQALHGR
	orf15-1	10 20 30 40 50 60	MQARLLIPIILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAANKMDLQALHGR
40	orf15a.pep	70 80 90 100 110 120	KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
	orf15-1	70 80 90 100 110 120	KVALYIATMGDQSGSGLTGGRYSIDALIRGEYINSPAVRTDYTPRYETTAETTSGLTG
45	orf15a.pep	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15-1	130 140 150 160 170 180	LTSLSTLNAPALSRTQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
50	orf15a.pep	190 200 210 220 230 240	FLRGIDVVPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15-1	190 200 210 220 230 240	FLRGIDVVPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
55	orf15a.pep	250 260 270 280 290 300	IKPKNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS DIQPYGNHMGNSAPSVEADN
	orf15-1	250 260 270 280 290 300	IKPKNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFS DIRPYGNHTGNSAPSVEADN
60	orf15a.pep	310 320	SHEGYGSDEAVRRHRQGPX
	orf15-1		SHEGYGSDEVVRQHRQGPX

310

320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```

1  ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
5  51  CGCCTGCGGG ACACGTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
101 TCGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251 TTGATGCACT GATTTCGCGG GAATACATAA ACAGCCCTGC CGTCCGCACC
301 GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
10 351 TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
15 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAAAC TGAATATTTT GCAGTAGACA
701 GAACCAATAA AAAATTGCTC ATCAACCCA AAACCAATGC GTTTGAAGCT
751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801 AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTCTCC GATATCCAAC
20 851 CATACGCGAA TCATACGGGT AACTCCGCC CATCCGTA GAATGATAAC
901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAC AACATAGACA
951 AGGGCAACCT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```

1  MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAARK
25 51  DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGGLTG LTSLSTLNA PALSRQSDG SGSRSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSF ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
30 301 SHEGYGYSDE AVRQHRQGP *

```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```

35 orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAARKDMDLQALHGR 60
   orf15ng   MRARLLIPILFSVFILSACGTLTGIPSHGGKRFVEQELVAASARAARKDMDLQALHGR 60

   orf15.pep  KVALYIATMGDQSGSLTGGRYSIDAXXGEYINSPAVRTDYTYPRYETTAETTSGGLTG 120
   orf15ng    KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG 120

40 orf15.pep  LTSLSTLNAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF 180
   orf15ng    LTSLSTLNAPALSRQSDGSGSRSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF 180

45 orf15.pep  FLRGIDVVSFANADTDVFINIDVFGTIRNRTEM 213
   orf15ng    FLRGIDVVSFANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL 240

```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```

50 orf15-1.pep  10 20 30 40 50 60
   orf15ng     10 20 30 40 50 60
   orf15-1.pep  70 80 90 100 110 120
   orf15ng      70 80 90 100 110 120
   orf15-1.pep 130 140 150 160 170 180
   orf15ng      130 140 150 160 170 180

```

10

15

20

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

35

40

45

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

1 ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
51 AEKTLHTDPO TASRPLPGLP XLTAVSTLEG TMSSWVGIGG GSLSVPFLIH

101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV
151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

```

5      1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTGCGCGTAG GCAGTGCAGG
51     AGGTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101    CTGTCGTTTT ATGGGTGCTT GATTGTCAGG GTTTGGCACA ACATCCTTAC
151    GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201    CTTTTCAGT  ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251    CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
10     301 CCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351    GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
401    CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCCACTG
451    TTCGGCACAA TGTCGAGCTG GGTCCGCATA GCGGCGGTT CACTTCCGT
501    CCTCTTCTTA ATCCACTGCG GCTTCCCGC CCATAAAGCC ATCGGCACAT
15     551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601    CTCACCGGCC TGAATATTGC AGGATTGCCG GAAGGTCAC TGGGCTTCCT
651    TTACCTGCCG GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTGCCC
701    CGCTCGGTGT CAAAACGCC CACAACTTT CTTCTGCCAA ACTCAAAAAA
20     751 Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801    GCTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

```

25     1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
51     AQHLAVGTSF AVMVFTAESS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA
101    LSAKYIPAFG LQIFILFLT AVAFKTLHTD PQASRPLPG LPGLTAVSTL
151    FGTSSSWVGI GGGSLVFPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201    LNLNLIAGLP EGSGLFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251    XFGIMLLLIA GKMLYNLL*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

```

ORF17  3  HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
HK  +  +  V  +  P  ++  VF  G  F  +  +IF  +++L  ++  D
HI0902 72  HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMLVSIKKD- 130

35     ORF17  60  QTASRPLPGLPXLTA VSTLFGTMSWVGIGGSLVFPFLIHCGFPAHKAIGTSSGLAWPI 119
Q  ++  L  L  +  L  G  SS  GIGG  VPFL  G  +AIG+S+  +
HI0902 131  QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189

40     ORF17  120  ALSGAISYLLNGLNLIAGLPEGSLGFLYLPVAVLSAATIAFAPLGVXXXXXXXXXXXXXXXXX 179
+SG  S+++G  +PE  SLG++YLPVAV  ++A  +  +  LG
HI0902 190  GISGMFSFIVSGWGNPLMPEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKG 249

ORF17  180  FGIMLLLIAGKM 191
F  +  L+++A  M
45     HI0902 250  FALFLIVVAINM 261

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis*:

```

50     orf17.pep                                10      20      30
                                           GQHKQAVNGKTVFTMMPGMIFGVFTGAFS
                                           |||||: |||||: |||||: |||
orf17a  QGLAQHPYAQHLAVGTSFAVMVFTA FSSMLGQHKQAVDWKTVFTMMPGMVFGVFAGALS
                                           50      60      70      80      90     100

55     orf17.pep                                40      50      60      70      80      90
AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSWVGIGG
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
orf17a  AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPGLTA VSTLFGTMSWVGIGG

```

-103-

		110	120	130	140	150	160
5	orf17.pep	100	110	120	130	140	150
	orf17a	170	180	190	200	210	220
10	orf17.pep	160	170	180	190		
	orf17a	230	240	250	260		

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCCGTAG	GCAGTGC	GCGC
	51	AGGTTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC	
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC	
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTGATGG	TCTTCACCGC	
20	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA	
	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA	
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATTT	TCTTCATCCT	
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG	
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TCCACACTG	
25	451	TTCGGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTTCCGT	
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCGC	CCATAAAGCC	ATCGGCACAT	
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG	
	601	CTCAACGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT	
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC	
	701	CGCTCGGTG	CAAAACGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAAAA	
30	751	TCCTTCGGCA	TTATGTTGCT	TTGATTGCC	GGAAAAATGC	TGTACAACCT	
	801	GCTTTAA					

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGT	LIVPVVLWVL	DLQGLAQHPY	
35	51	AQLAVGTSF	AVMVFAFSS	MLGQHKQAV	DWKTFTMMP	GMVFGVFAGA	
	101	LSAKYIPAFG	LQIFFILELT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL	
	151	FGTMSSWVGI	GGGSLVPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL	
	201	LNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK	
	251	SFGIMLLLIA	GKMLYNLL*				

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40	orf17a.pep	10	20	30	40	50	60
	orf17-1	10	20	30	40	50	60
45	orf17a.pep	70	80	90	100	110	120
	orf17-1	70	80	90	100	110	120
50	orf17a.pep	130	140	150	160	170	180
	orf17-1	130	140	150	160	170	180
55	orf17a.pep	190	200	210	220	230	240
	orf17-1	190	200	210	220	230	240
60	orf17a.pep	250	260	269			
	orf17-1	250	260	269			
65	orf17a.pep						

-104-

orfl7-1 ||||| |||||
 HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
 250 260

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N.gonorrhoeae*:

	orfl7.pep	GQHKQAVNGKTVFTMMPGMIFGVFTGAFS	30
		: : : : :	
10	orfl7ng	QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKQAVDWKTI FAMMPGMIFGVFAGALS	102
	orfl7.pep	AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSWVGIGG	90
		: : : : :	
15	orfl7ng	AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGG	162
	orfl7.pep	GSLVSPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPV	150
		: : : : :	
	orfl7ng	GSLVSPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPV	202
20	orfl7.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFSGIMLLLIAGKMLYNLL	196
		: : : : :	
	orfl7ng	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFSGIMLLLIAGKMLYNLL	268

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

25	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVFTAFSS	MLGQHKQAV	DWKTI FAMMP	GMIFGVFAGA
	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTG	RQTASRPLPG	LPGLTAVSTL
	151	FGAMSSWVG	GGGSLVSPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	VNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKE
30	251	SFGIMLLLIA	GKMLYNLL*			

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

	1	ATGTGGCATT	GGGACATTAT	CTTAATCCTG	CTTGCcgtag	gcAGTGC GGC
	51	AGGTTTTATT	GCCGGCCTGT	Tcgggtgtagg	cggcgGTACG	CTGATTGTCC
35	101	CTGTCGTTTT	ATGGGTGCTT	GATTTGCAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CacaTccttc	gcCGTCATGG	TCTTCACCGC
	201	CTTTTCCAGT	ATGTTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
	251	CCATATTGTC	GATGATGCCG	GGTATGATAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	CGCGTTCGGG	CTTCAAATTT	TCTTCATCCT
	351	GTTTTTAAAC	GCCGTCGCAT	TCAAAACACT	GCATACCGGT	CGTCAGACGG
40	401	CATCCCGCCC	GCTGCCCGGG	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCCGGCGCAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACCTTCCGT
	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAAGCC	ATCGGCACAT
	551	CATCCGCGCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	GTCAACGGTC	TGAATATTGC	AGGATTGCCC	GAAGGTCGC	TGGGCTTCCT
45	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCCC
	701	CGCTCGGTGT	CAAAACCGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAGAA
	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

50	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAVGTSF	AVMVFTAFSS	MLGQHKQAV	DWKTI FAMMP	GMIFGVFAGA
	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTG	RQTASRPLPG	LPGLTAVSTL
	151	FGAMSSWVG	GGGSLVSPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	VNGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKE
55	251	SFGIMLLLIA	GKMLYNLL*			

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

	10	20	30	40	50	60
orfl7-1.pep	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPYAQHLAVGTSF	

-105-

5	orf17ng-1	 MWHWDIILILLAVGSAAGFIAGLFGVGGGTLLVPPVVLWVLDLQGLAQHPYAQHLAVGTSF
		10 20 30 40 50 60
10	orf17-1.pep	70 80 90 100 110 120 AVMVFTAFSSMLGQHKQAVDWKTIVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
	orf17ng-1	 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
15	orf17-1.pep	130 140 150 160 170 180 AVAFKTLHTDPQTASRLPGLPGLTAVSTLFGTMSSWVGIGGSLSVFPFLIHCGFPAHKA
	orf17ng-1	 AVAFKTLHTGRQTASRLPGLPGLTAVSTLFGTMSSWVGIGGSLSVFPFLIHCGFPAHKA
20	orf17-1.pep	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
	orf17ng-1	 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAAVAVLSAATIAFAPLGVKTA
25	orf17-1.pep	250 260 269 HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
	orf17ng-1	 HKLSSAKLKESFGIMLLLIAGKMLYNLLX
		250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

30	sp P44070 Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir G64015 hypothetical protein HI0902 - Haemophilus influenzae (strain Rd KW20) gi 1573922 (U32772) H. influenzae predicted coding region HI0902 [Haemophilus influenzae] length = 264 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 15/43 (34%), Positives = 23/43 (53%)	
	Query:	55 AVGTSFAVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVF 97 A+GTSEFA +V T S HK + W+ + + P ++ VF
35	Sbjct:	52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94
	Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23 Identities = 44/114 (38%), Positives = 65/114 (57%)	
40	Query:	150 LFGAMSSWVGIGGSLSVFPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209 L G SS GIGG VPFL G +AIG+S+ + +SG S++V+G +
	Sbjct:	148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCEGMLLGISGMFSFIVSGWGNPLM 207
45	Query:	210 PEGSLGFLYLPAAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263 PE SLG++YLPV ++A + + LG KL + LK+ F + L+++A M
	Sbjct:	208 PEYSLGYIYLPVAVLGITATSEFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

This analysis, including the homology with the hypothetical *H. influenzae* transmembrane protein, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

55 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 95>:

55	1	..GGAACGAT GGCAGGCAGA CCCCAGCAT CCGCTGCTCG GGCTTTTTCG
	51	CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCGGAATA TGTGCGTTGG
60	101	TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTGTGTGT TGCAGGCACTG
	151	CTCAAACCTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
	201	GCTGATGGCG GTTGCTATG TCCACCGCTG CGGTATAGAC CGGCAGCCGC
	251	CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGCGAGCG

-106-

301 TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
351 A

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

5 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
51 LKLYALKPVY WFLVQFVLM VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
101 LMQVSVLVLL LSEIGR*

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

10 1 ATGATTTTGC TGCATTGGGA TTTTTGTCT GCCTTACTGT ATGCGGCGGT
51 TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTAA
201 CCTGACTTTG GGCAGCATAT TTTTTTCAT CGGGCATTGG AACCGGAAAA
251 CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
15 351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGCGG
401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTCAG
451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG
20 551 CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
601 AGATAA

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

25 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLOW FWASIMLWLG ISVLGAKLMP
51 GIWGMTRAAP LFIPHFYLT LGSIIFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSM TLAFVICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLO
151 FVLMAYAVH RCGIDRQPPS TFGGSQRLRG GLTAALMQVS VLVLLLSEIG
201 R*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N.*

30 *meningitidis*:

10 20 30
orf18.pep GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
35 orf18a TRAAPLFIPHFYLTLSIIFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
60 70 80 90 100 110
40 40 50 60 70 80 90
orf18.pep CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLMAYAVVHRCGIDRQPPSTFGGS
40 orf18a CALVHYCFSTVQVFVFAALLKLYALKPVYWFVLMAYAVVHRCGIDRQPPSTFGGS
120 130 140 150 160 170
45 100 110
orf18.pep QLRLGGLTAALMQVSVLVLLSEIGRX
45 orf18a QLRLGGLTAALMQVSVLVLLSEIGRX
180 190 200

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

50 1 ATGATTTTGC TGCATTGGGA TTTTTGTCT GCCTTACTGT ATGCGGCGGT
51 TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151 GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTTCATCC CCCATTTTAA
201 CCTGACTTTG GGCAGCATAT TTTTTTCAT CGGGCATTGG AACCGGAAAA
251 CGGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCTCT GCTCGGGCTG
55 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCT GAATATGTGC
351 GTTGGTGCAT TATTGCTTTT CGNGAACGGT TCAAGTGTTT GTGTTTGCGG
401 CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTCAG

-107-

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
 501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG
 551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
 601 AGATAA

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
 51 GIWGMTRAAP LFIPHFYLT LGSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSM TLAFV GICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ
 151 FVLMAYVYH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLLSEIG
 201 R*

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

		10	20	30	40	50	60
15	orf18a.pep	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP	GIWGMTRAAP
	orf18-1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLG	ISVLGAKLMP	GIWGMTRAAP
		10	20	30	40	50	60
20	orf18a.pep	LFIPHFYLT	LGSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFV GICALVH
	orf18-1	LFIPHFYLT	LGSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSM TL	AFV GICALVH
		70	80	90	100	110	120
25	orf18a.pep	YCFSXTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAYVYH	RCGIDRQPPS	TFGGSQLRLG
	orf18-1	YCFSXTVQVF	VFAALLKLYA	LKPVYWFVLQ	FVLMAYVYH	RCGIDRQPPS	TFGGSQLRLG
		130	140	150	160	170	180
30	orf18a.pep	GLTAALMQXS	VLVLLLSEIG	RX			
	orf18-1	GLTAALMQXS	VLVLLLSEIG	RX			
		190	200				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N.gonorrhoeae*:

40	orf18.pep		GNGWQADPEHPLLGLFAVSNVSM TLAFVGI	30
	orf18ng	TRAAPLFIPHFYLT	LGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGI	115
45	orf18.pep	CALVHYCFSGTVQVF	VFAALLKLYA LKPVYWFVLQ FVLMAYVYHRCGIDRQPPSTFGGS	90
	orf18ng	CALVHYCFSGTVQVF	VFAALLKLYA LKPVYWFVLQ FVLMAYVYHRCGIDRQPPSTFGGS	175
	orf18.pep	QLRLGGLTAALMQXS	VLVLLLSEIGR	116
	orf18ng	QLRLGVLAAML	MQVAVTAMLLAEIGR	201

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

1 ATGATTTTGC TGCATTGGA TTTTGTCT GCCTTACTGT aTGC GGcgg
 51 tttTctgTTT CTGATATCC GCGCAGGAAT GTTGCAATGG TTTGGGCGA
 101 GTATTGCGTT GTGGCTCGGC ATCTCGTTT TAGGGGTAA GCTGATGCCG
 151 GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTTCATCC CCCATTTTA
 201 CCTGACTTTG GGCAGCATAT TTTTTCAT CGGGTATTGG AACCGGAAAA
 251 CAGATGAAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGCTT
 301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTGC GAATATGTGC
 351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTC GTGTTTGGCG
 401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGTTT CGTGTTCAG
 451 TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
 501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGCGCG

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLGI ISVLGVKLMP
5 51 GMWGMTRAAP LFIPHFYLTLS GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSM TL AFGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAYAVVH RCGIDRQPPS TFGGSQRLRG VLAAMLMOVA VTAMLLAEIG
201 R*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10		10	20	30	40	50	60
	orf18-1.pep	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP					
	orf18ng	MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIALWLGISVLGVKLMPGMWGMTRAAP					
15		10	20	30	40	50	60
	orf18-1.pep	70	80	90	100	110	120
	orf18ng	LFIPHFYLTLSGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSM TLAFVGICALVH					
20		70	80	90	100	110	120
	orf18-1.pep	130	140	150	160	170	180
	orf18ng	YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQRLRG					
25		130	140	150	160	170	180
	orf18-1.pep	190	200				
	orf18ng	GLTAALMQVSVLVLLSEIGRX					
30		190	200				
	orf18-1.pep	VLAAMLMOVAVTAMLLAEIGRX					
	orf18ng						

Based on this analysis, including the presence of several putative transmembrane domains in the
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
201 CCTGTTTACC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC
251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
45 301 GGCGCGGNCG ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNXXTGRLK NIITVALFT LSSLTAQSTL GTGLPFILAM TLMTXXFTIL
101 GAX...

50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:

1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

201 CCTGTTCCACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
 301 GGC CGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 5 401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
 451 CTGTTCCAAA TCGTCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
 551 ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCTGT
 10 651 TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701 GTTACTACTT TGCCGCCCAA GACATACAG AACGCATCAG CTCGCCCCAC
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
 801 CCGCATCCAC CGCTGCTCG AAATGCAGG ACAAGCCTGC CGCAACACCG
 851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 15 901 CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951 CGACAGTCCC GACATCCGCC ACCTGCGCGG CCTTCTCGAC AACCTCGGCA
 1001 GCCTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
 1101 CAAAAACCC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
 20 1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
 1201 ATCGTCGAAG CCTCAACCT CAACTCGGC TACTGGATAC TACTGACCGC
 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTCCGCC
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 25 1351 TACTTCACCC CGTCTGTCGA AACCAAATC TGGATTGTCA TCGCCAGTAC
 1401 CACCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCTCGCAGG TTTGGACGTA
 1501 TACGCCGCCA TGCCCGTAGC CATCATCGAC ACCATTATCG GCGCATCCCT
 1551 TGCTGGGCG GCAGTCAGT ACCTGTGGCC AGACTGGAAA TACCTCACGC
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
 30 1651 AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
 1701 CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
 1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
 1801 CCGGGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
 35 1851 CGCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
 1901 TTACCGCACA GTTCCACCTC GCGCCGAAC ACACCGCCA CATCTTCAA
 1951 CACCTGCCCG AAACCGAACC CGACGACTT CAGACAGCAC TGGATACACT
 2001 GCGCGCGGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCT CCAACAGCTC CAACTCATCG CCGACAGCT CGAACCTAC
 40 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAG CAGCCCCAAA ATGCAGCCTG
 2151 A

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

1 MKTPLLKPLL ITS LPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGR LK NIITVALFT LSSLTAQSTL GTGLPFILAM TLMTEGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
 45 151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
 201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIIFRIH RLLEMGGQAC RNTAQALRAS KDYVYSKRLG
 301 RAIEGCRQSL RLLSDSNDSP DIRHLRRLD NLGSVDQQFR QLQHNGLQAE
 351 NDRMGDTRIA ALETSSLKNT WQAIRPOLNL ESGVFRHAVR LSLVVAAGT
 50 401 IVEALNINLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
 451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE
 551 KITERLKS GE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
 601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
 55 651 HLPETEPDDF QALDTRLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
 701 YRAYRQIPHR QPQNAA*

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60 orf19 6 LKPLLITSLP VFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD LDNXXTGR LKNIIT 65
 L +I+++PVF +V AA +W +MP +LGIIAGGLVD LDN TGR LKN+ T
 YHFK 5 LNAKVISTIP VFIAVNIAAVGIWFFDISSQSMPLILGIIAGGLVD LDNRLTGR LKNVFT 64

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N*.

meningitidis:

		10	20	30	40	50	60
10	orf19.pep	<u>MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGI</u> IAGGLVDLNXXTGRLK					
	orf19a	<u>MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGI</u> IAGGLVDLNRLTGRLK					
		10	20	30	40	50	60
15		70	80	90	100		
	orf19.pep	NIITTVALFTLSSSLTAQSTLGTGLPFI <u>LAMTLM</u> TXXTILGAX					
	orf19a	NIIATVALFTLSSSLVAQSTLGTGLPFI <u>LAMTLM</u> FGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
20							
	orf19a	TTLTYTPETYWLTP <u>FMIL</u> CGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

	1	ATGAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
25	51	CGCCAGTGTC	TTTACGCGCG	CCTCCATCGT	CTGGCAGCTG	GGCGAACCCA
	101	AGCTCGCCAT	GCCTTTCGTA	CTCGGCATCA	TCGCTGGCGG	CTCGGTGCGT
	151	TTGGACAACC	GCCCTGACCG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
	201	CCTGTTACAC	CTCTCCTCAC	TTGTGCGGCA	AAGCACCCCT	GGCACAGGTT
30	251	TGCCATTATC	CCTCGCCATG	ACCCTGATGA	CTTTTCGGCTT	TACCATCATG
	301	GGCGCGGTCT	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACCTCGCGT
	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC	TGGCTTGACC
	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC	CGCCATCATC
35	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA	AGCTCGCCAA
	501	CGCCTACGAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC	ACGTTTTTCG
	551	ATCCCAGACG	AGCCGAATGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
	601	AGCAACACCG	CGGTATCATC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
40	651	TTACCGCCTT	CGCGCAAAC	ACCGCCACCC	CGCGACCGGC	AAAATGTGTG
	701	GCTACTACTT	CGCGGCCCAA	GACATACACG	AACGCATCAG	CTCGGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTCG	CGCAACACCG
45	851	CCCAAGCCCT	CGCGCGAAG	AAAGACTACG	TTTACGCAA	ACGCTCTCGG
	901	CGCGCCATCG	AAGGCTGCCG	CCAATCGCTG	CGCCTCTTTT	CAGACAGCAA
	951	CGACAATCCC	GACATCGGCC	ACCTGCGCCG	CCTTCTCGAC	AACCTCGGCA
	1001	CGCTCGACCA	CGAGTTCCCG	CAACTCCAGC	ACAACGGCCT	CGAGGCAGAA
50	1051	AACGACCGCA	TGGGCGACAC	CCGCATCGCC	CGCCTCGAAA	CCGGACCGCT
	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC	GAATCAGGCG
	1151	TATTCGCGCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCGCG	CGCCTGCACC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGCG	TACTGGATAC	TACTGACCGC
55	1251	CCTTTTCGTC	TGCCAACCCA	ACTACACCGC	CACCAAAAGC	CGCGCTCGCC
	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTCGGGTC	GCTCGTCCCC
	1351	TACTTTTACC	CTCTCGTCGA	AACCAAATCT	TGGATCGTCA	TCGCCAGTAC
	1401	CACCCCTTTT	TTCATGACCC	GCACCTACAA	ATACAGCTTC	TCGACATTTT
60	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG	GTTGGACGTA
	1501	TACGCGGCCA	TGCCCGTACG	CATCATCGAC	ACCATTATCG	GCGCATCCCT
	1551	TGCCTGGGCG	CGAGTCAGCT	ACCTGTGGCC	AGACTGGAAA	TACCTCACGC
	1601	TCGAACGCAC	CGCGCCCTTT	GCCGTATGCA	CGAACGGCGC	CTATCTCGAA
65	1651	AAAATCACCG	AACGCCTCAA	AAGCGGCGAA	ACCGGCGACG	ACGTCGAATA
	1701	CCGCGCCACC	CGCGGCCGCG	CCCACGAACA	CACCGCGCCG	CTCAGCAGCA
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCCGCAA	AATTCGCCGA	CAGCTTGCAA
	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGCTAC	CGCCTGACCG	GCTACATCTC
70	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC	AGCCCCGACT
	1901	TTACCGCACA	GTTCCACCTC	GCCGCCGAAC	ACACGGCCCA	CATCTTCCAA
	1951	CACCTGCCCC	AAACCGAACC	CGACGACTTT	CAGACAGCAC	TGATACACAT
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA	ACACAAAGCC
75	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGAGCT	CGAACCCTAC
	2101	TACCGCGCCT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA	ACGCAGCCTG
	2151	A				

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

1  MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51  LDNRLTGRLL NIIATVALEF LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI
151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYFFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMGGQAC RNTAQLRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSDNDP DIRHLRRLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAGT
401 IVEALNINLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQLT TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAHTAHIFQ
651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIRQLEPY
701 YRAYRQIPHR QPQNAA*

```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

		10	20	30	40	50	60
20	orf19a.pep	MKTPPLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL
	orf19-1	MKTPPLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL
		10	20	30	40	50	60
25	orf19a.pep	NIIATVALEF	LSSLVAQSTL	GTGLPFILAM	TLMTFGFTIM	GAVGLKYRTF	AFGALAVATY
	orf19-1	NIITTVALEF	LSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL	GAVGLKYRTF	AFGALAVATY
		70	80	90	100	110	120
30	orf19a.pep	TTLTYTPETY	WLTNPFMILC	GTVLYSTAIL	LFQIILPHRP	VQENVANAYE	ALGSYLEAKA
	orf19-1	TTLTYTPETY	WLTNPFMILC	GTVLYSTAIL	LFQIVLPHRP	VQESVANAYD	DALGSYLEAKA
		130	140	150	160	170	180
35	orf19a.pep	DFFDPDEAEW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKRRHPRTA	KMLRYFFAAQ
	orf19-1	DFFDPDEAAW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKRRHPRTA	KMLRYFFAAQ
		190	200	210	220	230	240
40	orf19a.pep	DIHERISSAH	VQEMSEKFKN	TDIIFRIHRL	LEMGGQACRN	TAQALRASKD	YVYSKRLG
	orf19-1	DIHERISSAH	VQEMSEKFKN	TDIIFRIHRL	LEMGGQACRN	TAQALRASKD	YVYSKRLG
		250	260	270	280	290	300
45	orf19a.pep	RAIEGCRQSL	RLLSDSDNDP	DIRHLRRLD	NLGSVDQQFR	QLQHNGLQAE	ENDRMGDTRIA
	orf19-1	RAIEGCRQSL	RLLSDSDNDP	DIRHLRRLD	NLGSVDQQFR	QLQHNGLQAE	ENDRMGDTRIA
		310	320	330	340	350	360
50	orf19a.pep	ALETGSLKNT	WQAIRPQLNL	ESGVFRHAVR	LSLVVAAGT	IVEALNINLG	YWILLTALFV
	orf19-1	ALETSSLKNT	WQAIRPQLNL	ESGVFRHAVR	LSLVVAAGT	IVEALNINLG	YWILLTALFV
		370	380	390	400	410	420
55	orf19a.pep	CQPNYTATKS	RVQRRIAGTV	LGIVIVGSLV	PYFTPSVETK	LWIVIASTTL	FFMTRTYKYSF
	orf19-1	CQPNYTATKS	RVQRRIAGTV	LGIVIVGSLV	PYFTPSVETK	LWIVIASTTL	FFMTRTYKYSF
		430	440	450	460	470	480
60	orf19a.pep	STFFITIQLT	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTLERTAAL
	orf19-1	STFFITIQLT	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTLERTAAL
		490	500	510	520	530	540
65	orf19a.pep	STFFITIQLT	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTLERTAAL
	orf19-1	STFFITIQLT	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTLERTAAL

-112-

		STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL	
		490 500 510 520 530 540	
5	orf19-1	550 560 570 580 590 600	
	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	
10	orf19-1	550 560 570 580 590 600	
	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHAAEHTAHIFQHLPETEPDDF	
15	orf19-1	610 620 630 640 650 660	
	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHAAEHTAHIFQHLPETEPDDF	
20	orf19-1	670 680 690 700 710	
	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHRQPQNAAX	
	orf19-1	670 680 690 700 710	
	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLIARQLEPYRAYRQIPHRQPQNAAX	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N.*

gonorrhoeae:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng		
30	orf19.pep	NIIITVALEFTLSSSLTAQSTLGTGLPFILAMTLMXXFTILGAX	103
	orf19ng		
	orf19ng	NIIATVALEFTLSSSLTAQSTLGTGLPFILAMTLMFTGFTILGAVGLKYRTFAFGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNRLTGRLE	NIIATVALEFT	LSSSLTAQSTL	GTGLPFILAM	TLMTFGFTIL
40	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAI
	151	LFQIILPHRP	VQESVANAYE	ALGGYLEAKA	DDFDPDEAAW	IGNRHIDLAM
	201	SDTGVTAFN	QCRSALFYRL	RGKRRHPRTA	KMLRYFFAAQ	DIHERISSAH
	251	VQYQEMSEKF	KNTDIIIFRIR	RLLEMGGQAC	RNTAQAIRSG	KDYVYSKRLG
	301	RAIEGCRQSL	RLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHS DSPAE
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGCGCG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
50	201	CCTGTTTACC	CTCTCCTCGC	TCACGCGCGA	AAGCACCCCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
	301	GGCGCGGTG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
55	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCTGCCC	CCACCGCCCC	GTCCAAGAAA	GCGTCGCCAA
	501	TGCTTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCGATGA	GGCAGCTGG	ATAGGCAACC	GCCACATCGA	CCTCGCCATG
60	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAA	ACCGCAACCC	GCGCACCGCC	AAAATGCTGC
	701	GCTACTACTT	CGCGGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAATTC	AAAAACACCG	ACATCATCTT
	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGTCGGGG	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
	901	CGCGCCATcg	aaggctgCCG	CCAGTCGCTg	cgccctCCTTt	cagacggcaA
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGca

-113-

5
10
15
20
25

```

1001 GCGTcgacca gcagtTCgc caactCCGAC ACAGcgactC CCCCgcgaa
1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCctgaaa ccggcagctT
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCTTGCACC
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CCTCCGTCGA AACCAAATC TGGATTGTCA TCGCCGGTAC
1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTCGC CATCATcgaC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCGGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
1651 AAAATTGCCG AACGCCTCAA AACCGCGGAA ACCGGCGACG ACATAGAATA
1701 CACGCATACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTGCGCGA CAGCCTGCAA
1801 CCGGGCTTTA CCCTGCTCAA AACCGGTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
2001 GCGCGCGGAA CTCGGCACCC TCCGCACCG CAGCAGCGGA ACACAAAGCC
2051 ACATCTCCT CCAACAGCTC CAACTCATCG CcggGCAACT CGAACCTTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A

```

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

30
35
40

```

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNRLTGRLL NIIATVALFT LSSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101 GAVGLKYRTF AFGALAVATY TLTYPETYPETY WLTNPFMILC GTVLYSTAIL
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSDVDQFR QLRHSDSPA
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAACCT
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RYVQRIAGTV LGVIVGSLVP
451 YETPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITTIQAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSSGYLQ
551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPDMGPDDF QTALDTRLGE LGTLRTRSSG TQSHILLQOL QLIRQLEPY
701 YRAIRQIPHR QPQNAA*

```

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

45
50
55
60
65

```

          10      20      30      40      50      60
orf19-1.pep MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLL
          10      20      30      40      50      60
orf19ng-1   MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLL

          70      80      90      100     110     120
orf19-1.pep NIITTVALFTLSSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
          70      80      90      100     110     120
orf19ng-1   NIIATVALFTLSSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY

          130     140     150     160     170     180
orf19-1.pep TLTYPETYPETYWLTNPFMILCGTVLYSTAILLFQIIVLPHRPVQESVANAYDALGGYLEAKA
          130     140     150     160     170     180
orf19ng-1   TLTYPETYPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA

          190     200     210     220     230     240
orf19-1.pep DFFDPDEAAWIGNRHLIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYFFAAQ
          190     200     210     220     230     240
orf19ng-1   DFFDPDEAAWIGNRHLIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYFFAAQ

          250     260     270     280     290     300
orf19-1.pep DIHERISSAHVDYQEMSEKFKNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG

```

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

50
55
60
65

Query:	307	RQSLRLLSDGNDSPDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS	366
Sbjct:	1	RQSLRLLSDGNDSDIRHLSRLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS	60
Query:	367	FKNTWQAIRPQLNLESCVFRHAVRLSLVVAACTIVEALNNLNGYWILLTALFVCQPNYT	426
Sbjct:	61	FKNTWQAIRPQLNLESGVFRHAVRLSLVVAACTIVEALNNLNGYWILLTRLFVCQPNYT	120
Query:	427	ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT	486
Sbjct:	121	ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT	180
Query:	487	IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALVCSSG	546
Sbjct:	181	IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTERTAAALVCSSG	240

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADSLQPGFTLL 606
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFAD+ P
 Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSSTLSDMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632
 K ALTGYISALG ++ + +P
 Sbjct: 301 KPATALTYISALGHATAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15 1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
101 CGGGTATGCG GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG
151 CTTCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
201 TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAAGCCTTTA
251 TCCGCCATGT GCGCGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
351 TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTG GCTGCGGATT
401 ACGTTTCCTT ATATATTATT GATTTCCTG TCTTCATTG TCGGCTCGGT
451 ACTCAATTCT TATCATAAGT TCGGCATTCC GCGGTTTACG CCAC.GTTTC
25 501 TGAACGTGTC GTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
551 CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTGCGCG GCATTTTGCA
601 ACTCGrmtTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
651 CCAAACTGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
701 GCGCCTGCgA TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
30 751 CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
801 ACGCCGACCG CATGATGGAG CTGCCAGCG GCGTGCTGGG GCGGCGACTC
851 GGTACGATT TGTGCGCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
901 GGAACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGctgc
951 TGACGCTGCC GCGGgcGGTC GGACTGGCGG TGTGTCTGTT cCCgCtGGTG
35 1001 GCGACGCTGT TTATGTACCG CGWATTTACG CTGTTTGACG CGCAGATGAC
1051 GCAACACGCG CTGATTGCCT ATTCTTTTCG TTTAATCGGC TTAATCATGA
1101 TTAAGTGTTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAaWAmGCCC
1151 GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
40 1201 CTTTAyCGGC CCACTrrAAC rCaTTCGGAC TTTTCGCTTG CATCGGTCTG
1251 GCGCGGTGTA TCAATGCCGG ATTGTGTGTT TACCTGTTGC GCAGACACGG
1301 TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCCTT.AG CAAAAATGCT
1351 GcTCTCGCTC GCCGTGA

```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45 1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
101 LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151 LNSYHKFGIP AFTPKFLNVS FIVFALFFVP YFDPVVTAXA WAVFVGILQ
201 LXFLPLWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
50 301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL APGFYARONI XXPVKIAIFT LICKQLMNLX
401 FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
451 SRSP*

```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55 1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG

```

101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACT GCCCAACCTG
 151 CTTCCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TACAAGGAAA CGCGTCAAA AGAGGCGGCG GAGGCTTTTA
 251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
 5 301 CTGGGCATAC TTGCGCGGCC TTGGGTGATT TATGTTCCG CACCCGTTT
 351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA
 401 CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCATTTGT CGGCTCGGTA
 451 CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCTGCGCG TATTTTCGATC
 10 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTGCGCGG CATTTCGCAA
 601 CGTCGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 651 CAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701 CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
 15 801 CGCCGACCGC ATGATGGAGC TGCCGACGGG CGTGCTGGGG GCGGCACCTG
 851 GTACGATTTT GCTGCCGACT TTGTCAAAC ACTCGGCAA CCAAGATACG
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
 951 GACGCTGCCG CGGCGGTGCG GACTGGCGGT GTTGTGCTTC CCGCTGCTGG
 1001 GCACGCTGTT TATGTACCGC GAATTACGC TGTTCGACGC GCAGATGACG
 20 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
 1101 TAAAGTGTG GCACCCGGCT TCTATGCGCG GCAAACATC AAAACGCCCG
 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
 1201 TTTATCGGCC CACTGAAACA CGTCGGAATT TCGCTTGCCA TCGGTCTGGG
 1251 CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
 25 1301 TTTACCAACC TGGCAAGGGT TGGGACGCGT TCTTAGCAA AATGCTGCTC
 1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
 1401 GTTTGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGG CAGCTCTGCA
 1451 TCCTGATTGC CGTCGGCGGC GACTGTATT TCGCATCACT GGCGGCTTTG
 1501 GGCTCCGTC CGCGCCATT CAAACGCGTG GAAAACGA

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLIVTA
 101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
 151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPTALA WAVFVGILQ
 35 201 LGFQLPWLAK LGFLKLPKLS FKDAVNRVM QMAPAILGV SVAQVSLVIN
 251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLTLP AAVGLAVLSF PLVATLEMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNL
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 40 451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAL
 501 GFRPRHFKRV EN*

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45 Orf20 1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL LRRVFAEGAF 60
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDA FFVAFKLPNL LRR+FAEGAF
 MviN 14 MNLLKSLA AVSSMTMF SRVLGFARDAIVARI FGAGMATDA FFVAFKLPNL LRRIFAEGAF 73
 Orf20 61 AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLIVTALGILAAPWVIYVSAPSFAQD 120
 +QAQFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA
 50 MviN 74 SQAFVPILAEYKSKQGEATRIFVAVYSGLLTALAVVTVAGMLAAPWVIMVTAPGFADT 133
 Orf20 121 ADKFQLSIDLLRITFPYILLISLSSFVGSV LNSYHKFGIPAFTPXFLNVSFIVFALFFVP 180
 ADKF L+ LLRITFPYILLISL+S VG++LN+++F IPAF P FLN+S I FALF P
 55 MviN 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193
 Orf20 181 YFDPPTAXAWAVFVGILQLXFLQPLWLAKLGFLKLPKLSFKDAVNRVMQMAPAILGV 240
 YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV
 MviN 194 YFNPPVLALAWAVTVGGVLQVLYQLPYLKKIGMLVLPRIINFRDTGAMRVVKQMGPAILGV 253
 60 Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGV LG AALGTILLPTLSKHSANQDT 300
 SV+Q+SL+INTIFAS+L SGSVSWMYADR+ME PSGVLG ALGTILLP+LSK A+ +
 MviN 254 SVSISLIINTIFASFLASGSVSWMYADRLMEFSGVLGVALGTILLPSLSKSFASGNH 313

-117-

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G
 MviN 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYKFTAFDAAMTQRALIAYSFG 373

5 Orf20 361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXXCI 420
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPQG 440
 NA LL++ LR+ I+ P G
 MviN 434 NASLLYWQLRKQNIPTPQG 453

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf20.pep		MNMLGALAKVGS	SLTMVSRVLGF	VRDVTIARAF	GAGMATDAFF	VAFKLPNLL	RRVFAEGAF
orf20a		MNMLGALVKVGS	SLTMVSRVLGF	VRDVTIARAF	GAGMATDAFF	VAFKLPNLL	RRVFAEGAF
		10	20	30	40	50	60
orf20.pep		70	80	90	100	110	120
orf20a		AQAFVPILA	EYKETSKE	AXEAFIRH	VAGMLS	SFVLVIV	TALGILAAP
		70	80	90	100	110	120
orf20.pep		130	140	150	160	170	180
orf20a		ADKFQLS	IDLLRIT	TFPYILL	ISLSFV	GSVLNS	YHKFGI
		130	140	150	160	170	180
orf20.pep		190	200	210	220	230	240
orf20a		YFDPPTA	XAWAVF	VGGILQ	LFQLPW	LAKLGL	KLPKLS
		190	200	210	220	230	240
orf20.pep		250	260	270	280	290	300
orf20a		SVAQVSL	VINTIF	ASYLQ	SGSVS	WMYA	ADRM
		250	260	270	280	290	300
orf20.pep		310	320	330	340	350	360
orf20a		EQFSALL	DWGLRL	CMLLT	LPAAV	GLAVLS	FPLVAT
		310	320	330	340	350	360
orf20.pep		370	380	390	400	410	420
orf20a		LIGLIM	IKVLAP	GFYARQ	NIXXP	VKIAI	FTLIC
		370	380	390	400	410	420
orf20.pep		430	440	450			
orf20a		NAGLLF	YLLRRH	GIYQXP	QGLSV	LXQCC	SRSPX
		430	440	450	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

65 1 ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTTG GGATTGTGC GCGATACGGT CATTGCGCGC GCATTGGCGG
 101 CAGGCATGGC GACGATGCG TTCTTTGTGC CGTTCAAACT GCCCAACCTG

-118-

151 CTTCCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
 251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTCAT CGTTACCGCG
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTCCG CACCCGGTTT
 5 351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTG CTGCGGATTA
 401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTGT CGGCTCGGTA
 451 CTCAATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTTCCT
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTTCGATC
 10 551 CTCCCCTTAC CGCGCTGGCT TGGGCGGTTT TTGTCCGCGG CATTTCGCAA
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701 GACCTGCGAT TTTGGGCGTG AGCGTGCGCG AGATTTCTTT GGTGATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGCG AGCGTTTCAT GGATGTATTA
 801 CGCCGACCGC ATGATGGAAC TGCCCGGCGG CGTGCTGGGG GCGGCACTCG
 15 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGTCG GAATGGCGGT GTTGTGCTTC CCGCTGGTGG
 1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTTCGACGC GCAGATGACG
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
 20 1101 TAAAGTGTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
 1151 TCAAAATCGC CATCTTCACG CTCATTGCA CGCAGTTGAT GAACCTTGCC
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGCTGCGG
 1251 CGCGTGTATC AATGCCGAT TGTGTTTTC CCGTGTGCGC AGACACGGTA
 1301 TTTACCAACC TGGCAAGGTT TGGGACGCGT TCTTGCCAAA AATGCTGCTC
 25 1351 TCGCTCCCGC TGATGGGAGG CGGCCTGTAT GCCGCCCAA TCTGGCTGCC
 1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGCGGCC CGGCTCTCA
 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGGGCTTTG
 1501 GGCTTCGCTC CGCGCCATTT CAAACGCGTG GAAAGCTGA

This encodes a protein having amino acid sequence <SEQ ID 118>:

1 MNMLGALVKV GSLTMVSRVL GFVRDTVIAF AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPI LAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA
 101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
 151 LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPPTALAV WAVFVGGLIQ
 201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 35 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRXCMLLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 451 SLAVMGGGLY AAQIWLFPDW AHAGGMQKAA RLFILIAVGG GLYFASLAL
 40 501 GFRPRHFKRV ES*

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

		10	20	30	40	50	60
	orf20a.pep	MNMLGALVKV	GSLTMVSRVL	GFVRDTVIAF	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
45	orf20-1	MNMLGALAKV	GSLTMVSRVL	GFVRDTVIAF	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
		10	20	30	40	50	60
	orf20a.pep	70	80	90	100	110	120
50	orf20-1	AQAFVPI LAE	YKETRSKEA	EAFIRHVAGM	LSFVLVIVTA	LGILAAPWVI	YVSAPGFAKD
		70	80	90	100	110	120
	orf20a.pep	130	140	150	160	170	180
55	orf20-1	ADKFQLSIDL	LRITFPYILL	ISLSSFVGSV	LNSYHKFSIP	AFPTPTFLNVS	FIVFALFFVP
		130	140	150	160	170	180
	orf20a.pep	190	200	210	220	230	240
60	orf20-1	YFDPPTALAV	WAVFVGGLI	QLGFLPWLAK	LGFLKLPKLS	FKDAAVNRVM	KQMAPAILGV
		190	200	210	220	230	240
65	orf20a.pep	250	260	270	280	290	300
	orf20-1	SVAQISLVIN	TIFASYLQSG	SVSWMYADR	MMELPGGVLG	AALGTILLPT	LSKHSANQDT

ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae*:

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

-120-

1 MNMLGALAKV GSLTMVSRVL GFVRDVTIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPIAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YEDPPVTALA WAVFVGGILO
 5 201 LGFQLPWLA LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLFR KHGIYRPGQG LGQPSWRKCC
 10 451 SRSP*

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTTG GGATTGTGTC GCGATACGGT CATTGCGCGG GCATTGCGCG
 101 CCGGTATGGC GACGGATGCG TTTTTGTGCG CGTTCAAACG GCCCAACCTG
 15 151 CTTCCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCATA AGAGGCGAag gAGGCTTTTA
 251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG
 301 CTGGGCATAC TTGCCGCGcc tTGGGTGATT TATGTTtccg CgcccGGCTT
 351 TACCAAAGAC GCGGACAAGT TCCAACCTTC CATCAGCCTG CTGCGGATTA
 20 401 CGTTTTCTTA TATATATTG ATTTCTTTGT CTCTTTTGT CGGCTCGATA
 451 CTCAATTCTT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCTGTCCG TATTTTCGATC
 551 CGCCCCGTAC CGCGCTGGCG TGGGCGGTTT TTGTGCGCGG TATTTTGCAG
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 25 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CGCGCTCATG AAACAGATGG
 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttatCAAC
 751 ACGATTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
 801 cgCCGACCGC ATGATGGAGc tgcgcCGGGG CGTGTGGGG GCTGCACTCG
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 30 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCTGTT CCGCTGGTGG
 1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG
 1051 CAACACGCGC TGATTGCCTA TTCTTTCCGT TTAATCGGTT TAATTATGAT
 1101 TAAAGTGTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
 35 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
 1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG
 1251 CGCGTGCATC AACGCCGGAT TGTGTCTCTT CCTGTTGCGC AAACACGGTA
 1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
 1351 GCGCTCGCGG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC
 40 1401 GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
 1451 TCCTGATTGC CGTCGGCGGC GGAAGTGTATT TCGCATCTCT GGCGGCTTTG
 1501 GGCTTCGCTC CGGCCATTT CAAACGCGTG GAAAGCTGA

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

1 MNMLGALAKV GSLTMVSRVL GFVRDVTIAR AFGAGMATDA FFVAFKLPNL
 45 51 LRRVFAEGAF AQAFVPIAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
 101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVSGI
 151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YEDPPVTALA WAVFVGGILO
 201 LGFQLPWLA LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 50 251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL
 401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL
 451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
 501 GFRPRHFKRV ES*

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

	10	20	30	40	50	60
orf20-1.pep	MNMLGALAKV	GSLTMVSRVL	GFVRDVTIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
orf20ng-1	MNMLGALAKV	GSLTMVSRVL	GFVRDVTIAR	AFGAGMATDA	FFVAFKLPNL	LRRVFAEGAF
60	10	20	30	40	50	60
	70	80	90	100	110	120
orf20-1.pep	AQAFVPIAE	YKETRSKEA	EAFIRHVAG	MLSFVLIV	TALGILAA	PWVIYVSAPGFAQD
orf20ng-1	AQAFVPIAE	YKETRSKEA	EAFIRHVAG	MLSFVLIV	TALGILAA	PWVIYVSAPGFTKD
65						

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLRLITFPYILLISLSSSVGSLNSYHKFGIPAFPTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRLITFPYILLISLSSSVGSLNSYHKFGIPAFPTPTFLNISFIVFALFFVP					
		130	140	150	160	170	180
10	orf20-1.pep	YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
25		310	320	330	340	350	360
	orf20-1.pep	LIGLIMIKVLAPGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARONIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
30		370	380	390	400	410	420
	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRRKAG					
	orf20ng-1	NAGLLFFLLRKHGIRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRRKAG					
35		430	440	450	460	470	480
	orf20-1.pep	QLCILIAVGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAVGGGLYFASLAALGFRPRHFKRVESX					
40		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

45	sp P37169 MVIN_SALTY VIRULENCE FACTOR MVIN pir S40271 mviN protein - Salmonella typhimurium gi 438252 (Z26133) mviB gene product [Salmonella typhimurium] gn PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGSRLTMSRVLGFEVDRTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLLKSLAIVSSMTMF SRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILAIEYKETSKEATEAFIRHVAGMLSEVLIVVTALGILAAPWVIYVSAPGFTKD 120 +QAFVPILAIEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAFVPILAIEYKSKQGEATRIFVAYVSGLLTLALAVVTAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLLRLITFPYILLISLSSSVGSLNSYHKFGIPAFPTPTFLNISFIVFALFFVP 180 ADKF L+ LLRLITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALTTLRLITFPYILLISLASLVGAILNTWNRFSIPAFAPTPTFLNISMGIFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV Sbjct: 194 YFNPPVLALAWAVTVGGVLQVYQLPYLLKIGMLVLPRIINFRDGTAMRVVKQMGPAILGV 253
	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGVSWSMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFPSPGVLGVALGTILLPSLSKSFASGNH 313
70	

-122-

Query: 301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +ET FDA MTQ ALIAYS G
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYQKETAFDAAMTQRALIAYSVC 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
 LIGLI++KVLA GFY+RQ+IKTPVKIAI TLI TOLMNLAFIGPLKHAGLSL+IGL AC+
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIIYRPGRWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
 NA LL++ LRK I+ P GW VM L+ +P
 Sbjct: 434 NASLLYWQLRKQNIPTPQPGWWMFLMRLIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
 EW+ + + +L ++ G YFA+LA LGF+ + F R
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTcAC CGTGGCGAAA
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTcAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA tGGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEYNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
 45 201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTcAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTcAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTcATTAT
 501 CAAAGAAGCC GCCGAGGATT TCAACGCGCG CCGTGTGGTA TTGAGCCGTT
 551 TGACCGAAGC CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAATG CTGCCAATC CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 TGCCGGTTTG AGTGGCACGC ACATTcATT CATCGAGCCG CTCGGCGCGA
 701 ATAAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
 801 TTCTCAGTc AACAAACCGC GCCTCTTGGC TACCGTTTGG GGTGCGAAAG
 851 TATCGCAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
 901 TCCGTTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT

-123-

5
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACAACCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCAACACAGC
 1101 CGTCAACGGC GCGGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10
 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 15
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVITIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLLRVL GAKVSQITAG ELVDTDNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA POPDKYSITR
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRLKVL ETIEKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20
 1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CCGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201 GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
 25
 251 AGCGCGTACT TCACTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA CGCGCGGANGA
 351 ANTNNNGNNG AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401 GTCCGTTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCTG TGGTTGTGAT
 30
 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
 551 TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTGCGCG GCCCGCATCC
 651 GGCCGCTTTG AGTGGCACGC ACATTCAATT CATTGAGCCG GTCGGTGCAA
 701 ACAAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 35
 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801 TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 901 TCCGGTTTCG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951 GGGACGCTAC CACAATCAGA TTTCGTTAT CGAAGAAGGC CGCAGCAAAG
 40
 1001 AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
 1051 ACGACCTCG GCCATTTCTT GAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGT GCGGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA
 45
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
 1301 CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50
 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KXPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGXEXXX NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRRXLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHFAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLLRVL GAKVSQITAG ELVDADNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA POPDKYSITR
 55
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

60
 orf22.pep 10 20 30 40 50 60
 MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
 orf22a MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

-124-

		10	20	30	40	50	60
5	orf22.pep	70	80	90	100	110	120
	orf22a	70	80	90	100	110	120
10	orf22.pep	130	140	150			
	orf22a	130	140	150	160	170	180

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

15	orf22a.pep	10	20	30	40	50	60
	orf22-1	10	20	30	40	50	60
20	orf22a.pep	70	80	90	100	110	120
	orf22-1	70	80	90	100	110	120
25	orf22a.pep	130	140	150	160	170	180
	orf22-1	130	140	150	160	170	180
30	orf22a.pep	190	200	210	220	230	240
	orf22-1	190	200	210	220	230	240
35	orf22a.pep	250	260	270	280	290	300
	orf22-1	250	260	270	280	290	300
40	orf22a.pep	310	320	330	340	350	360
	orf22-1	310	320	330	340	350	360
45	orf22a.pep	370	380	390	400	410	420
	orf22-1	370	380	390	400	410	420
50	orf22a.pep	430	440				
	orf22-1	430	440				

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65	1	MIKIKKGLNL	PIAGRPEQVI	YDGPATEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 251 LFVTGRLNTE RVVALGGLQV NKPRLLRRTL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRY HN*

5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT
 10 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTCAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATGA GCAGCGAAAA
 351 AGTGCGCCGC AACCTGATT C AATCAGGCTT ATGGACTGCG CTTGCGACCC
 401 GTCCGTTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
 15 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATCAT
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
 551 TGACCGAAGC TAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCGG
 601 TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
 651 TGCCGGCTTG AGTGGCACGC ACATTCAAT CATCGAGCCA GTCGGCGCGA
 20 701 ATAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
 751 TTGTTCCGTA CAGGCCGTCT GAATACCGAG CGCGTGTTG CTTGGGCGG
 801 CCTGCAAGTC AACAAACCGC GCCTCTTGG TACCGTTTGG GGTGCGAAGG
 851 TGTCTCAACT TACCGCCGCG GAATTTGGTG ACGCGGACAA CCGCGTGATT
 901 TCCGGTTTCG TATTGAACGG TGCGATTGCA CAAGGCGCGC ATGATTATTT
 25 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGC
 1051 ACCACTCTCG GCCATTTCCCT AAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGC GCGACCGCG CCATGGTACC GATCGGCACT TATGAGCGCG
 1151 TAATGCCGTT GGACATCTCG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
 30 1201 GGCGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

1 MIKIKKGLNL PIAGRPEQVI YDGPATEVA LLGEEYVGM RPSMKIKEGEA
 35 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
 201 SENAANIETH EFGGPHAGL SGTHIHFIIEP VGANKTVWTI NYQDVIAIGR
 40 251 LFVTGRLNTE RVVALGGLQV NKPRLLRRTL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMLDIL PTLRLDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

45 overlap with ORF22ng:

orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
 orf22ng MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 60
 50 orf22.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYAPEALANLSGEEVRR 120
 orf22ng KKNPGVVFTAPASGKIAAIHRGEKRVLSVVIAVEGNDEIEFERYVPEALAKLSSEKVR 120
 orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
 55 orf22ng NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60 orf22-1.pep 10 20 30 40 50 60
 MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

	::: ::: :::
orf22ng-1	MIKIKKKGLNLPAGRPEQVIYDGPATTEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFE 10 20 30 40 50 60
	70 80 90 100 110 120
orf22-1.pep	KKNPGVVFTAPASGKIAATHRGEKRVLQS ^V VIAVEGNDEIEFERYAPEALANLSGEEVRR ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	KKNPGVVFTAPASGKIAATHRGEKRVLQS ^V VIAVEGNDEIEFERYVPEALAKLSSEKVR 70 80 90 100 110 120
	130 140 150 160 170 180
orf22-1.pep	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 130 140 150 160 170 180
	190 200 210 220 230 240
orf22-1.pep	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI 190 200 210 220 230 240
	250 260 270 280 290 300
orf22-1.pep	NYQDVITIGRLFATGRNLINTERVIALGGSQVNKPRLRLRTVLGAKVSQITAGELVDTDNRVI ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	NYQDVIAIGRLFVTGRNLINTERVVALGGLOVNKPRLRLRTVLGAKVSQITAGELVDADNRVI 250 260 270 280 290 300
	310 320 330 340 350 360
orf22-1.pep	SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELEFGWVAQPDPKYSITRTLGHFLKNK ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	SGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELEFGWVAQPDPKYSITRTLGHFLKNK 310 320 330 340 350 360
	370 380 390 400 410 420
orf22-1.pep	LKFNTAVNGGDRAMVP ^I GTYERVMPLDILPTLLLRDLIVGDTSDSAQALGCLELDEEDLA ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	LKFNTAVNGGDRAMVP ^I GTYERVMPLDILPTLLLRDLIVGDTSDSAQALGCLELDEEDLA 370 380 390 400 410 420
	430 440
orf22-1.pep	LCSFVCPGKYEYGPLL ^R KVLETIEKEGX ::: ::: ::: ::: ::: ::: ::: :::
orf22ng-1	LCSFVCPGKYEYGPLL ^R KVLETIEKEGX 430 440

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

orf22	1	MIKIKKGLNLPIAGRPEQAVYDGP	ITEVALLGEEYAGMRPSMKVKEGD	AVKKGQVLFED	60
48kDa	1	MI IKKGL+LPIAG P Q +++G +	EVA+LGEY GMRPSMKV+EGD	VKKGQVLFED	60
orf22	61	MITIKKGLDLPIAGTPAQVIHNG	NTVNEVAMLGEEYVGMRPSMKV	REGDVVKKGQVLFED	60
48kDa	61	KNKPGVVFTAPASGKIAAIHRGE	KRVLSQSVVIAVEXNDEIEFER	YAPEALANLSGEEVRR	120
orf22	121	KNKPGVVFTAPASG + I+RGE	KRVLSQSVVI VE +++I F RY	LA+LS E+V++	120
48kDa	121	KNKPGVVFTAPASGTVVTINRGE	KRVLSQSVVIKVEGDEQITF	TRYEAAQLASLSAEQVKQ	120
orf22	121	NLIQSGLWTALRTRPFSKIPAVD	AEPFAIFVNAMDTNP	158	
48kDa	121	NLI+SGLWTA RTRPFSPK+PA+DA	P +IFVNAMDTNP		
orf22	121	NLIESGLWTAFRTRPFSPKVPAL	DAIPSSIFVNAMDTNP	158	

Score = 530 bits (1351), Expect = e-150

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1 MIKIKKGLNLPAGRPQVIYDGPVITEVALLGEEYAGMRPXMVKVEGDAVKKGQVLFED 60
 5 Sbjet: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMPSMKVREGDVVKKGQVLFED 60

Query: 61 KKXPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXXX 120
 10 Sbjet: 61 KKNPGVVFTAPASGT VVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180
 15 Sbjet: 121 NLIESGLWTAFTTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 181 LSRL--TERKIHVCKAAGADVP--SENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTV 237
 20 Sbjet: 181 LTRLFNGQKPVYLC KDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRLINTERVIALGGSQVKNPRLRLTVLGAKVSQITAGELVDADN 297
 25 Sbjet: 241 WHLNYQDVIAIGKLF TGTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSVLNGAITQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 357
 30 Sbjet: 301 RVISGSVLNGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KKNLKFETTA VNGGDRAMVP IGT YERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 417
 35 Sbjet: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTDSAQNLGCLLEDEE 419

Query: 418 XXXXXSFVCPGKYEXG PLLRKVLETXEKEG 447
 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
 35 pleuropneumoniae] Length = 449
 Score = 555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27 MIKIKKGLNLPAGRPQVIYDGPATEVALLGEEYVGMPSMKIKEGEAVKKGQVLFED 86
 40 Sbjet: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146
 45 Sbjet: 61 KKNPGVVFTAPASGT VVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTV IIEAAEDFKRGLLV 206
 50 Sbjet: 121 NLIESGLWTAFTTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 207 LSRL--TERKIHVCKAAGADVP--SENAANIETHEFGGPHPAGLSGTHIHFIIEPVGANKTV 263
 55 Sbjet: 181 LTRLFNGQKPVYLC KDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRLINTERVVALGGLQVKNPRLRLTVLGAKVSQITAGELVDADN 323
 60 Sbjet: 241 WHLNYQDVIAIGKLF TGTGELFTDRIISLAGPQVKNPRLVTRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSVLNGAIAQGAHDYLG RYHNQISVIEEGRSKELFGWVAPQDPKYSITRTTLGHFL 383
 65 Sbjet: 301 RVISGSVLNGATAAGPVDYLG RYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KKNLKFETTA VNGGDRAMVP IGT YERVMXXXXXXXXXXXXXXXXXVGD TDSAQXXXXXXXXXX 443
 70 Sbjet: 361 K-KLFNFTTAVHGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTDSAQNLGCLLEDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
 ++VCPGK YGP+LR LE IEKEG
 Sbjet: 420 DLALCTYVCPGKNYGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

10 Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1  ..GCgNCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAAC TG
51  GTTTTTTATG GTAGCCAGTA CGTTGTGAT TGCTTTGATT GGTATTTTTG
101 TTA CTGAAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
151 TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251 TATGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20 401 GTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTTGsw CAkCATCTTT TTTGCCGCAC AGTTGTTCGC
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGCGCA
551 CGTTCCTTAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTG TATCGGTTTT
25 601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
651 ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGCGCGGCT
701 ACGGGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
801 GrkCmmTAC AAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTTCTTC TTGATTGCGT GGATTGCCTT ATTCTGCATT
30 901 TGGGTATTTg TTTTGGGCCT GCCGTCGGT CCGGCGCGC CCACATTCTA
951 TCCGACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
35 51  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101 TGLVSGSPFL KSIUVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301 WVEVLGLPVG PGAPTFYPAP *

```

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTCA
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151 GTCCCGATC CGCGCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
45 201 GATTACATT GTCAGCTGC TCAATGCCGA CGGTTTATC AAAATCCTGA
251 CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTG
301 GTTCTTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCGC
351 ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAACCT ACTACTTTA
50 401 TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
451 GTCGTCTTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501 TCCGCTTGCC GGTCTGGCTG CGGCTTTCG CGGCGTTTCG GGCGGTTATT

```

10

15

20

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

25

30

Computer analysis of this amino acid sequence gave the following results:

35

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N.*

meningitidis:

40

```

                                10      20      30
orf12.pep                      AXXIIHPXVVVGPEANWFFMVASTFVIALI
                               |   |||   ||||| ||||| ||||| |||||
orf12a    AAAFAGVSGGYSANLFLGTIDPLLAGITQAAQIIHDPDYVVGPEANWFFMVASTFVIALI
          180        190        200        210        220        230

```

45

orf12.pep
|
orf12a

40 50 60 70 80 90
GYFVTEKIVEPQLGPYQSDLSQEEDIRHSNEITPLEYKGLIWAGVVFFVALSALLAWSIV
|||
GYFVTEKIVEPQLGPYQSDLSQEEDIRHSNEITPLEYKGLIWAGVVFFVALSALLAWSIV
240 250 260 270 280 290

50

```

              100      110      120      130      140      150
orf12.pep  PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
           |||||
orf12a     PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMS
           |||||
              300      310      320      330      340      350

```

55

```

              160      170      180      190      200      210
orf12.pep  TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
            |||||
orf12a     TLGLYLVIIFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
              360      370      380      390      400      410

```

60

220 230 240 250 260 270
orf12.pep IGSASAQWAVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMSYFGLIMATVXXY

-130-

orf12a	IGSASAQWAVTAPIFVPMMLLAGYAPEVIAQAYRIGDSVTNIITPMMSYFGLIMATVIKY
	420 430 440 450 460 470
5	orf12.pep
	280 290 300 310 320
	KKDAGVGTLLIXMMLPYSAFFLIWIALFCIWVFLVGLPVGPGAPTFYPAPX
	orf12a
	KKDAGVGTLLISMMLPYSAFFLIWIALFCIWVFLVGLPVGPGAPTFYPAPX
	480 490 500 510 520

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

10	1	ATGAGTCAAA	CCGATACGCA	ACGGGACGGA	CGATTTTAC	GCACAGTCGA
	51	ATGGCTGGGC	AATATGTTGC	CGCACCCTGGT	TACGCTTTT	ATTATTTTCA
	101	TTGTGTTATT	GCTGATTGCC	TCTGCCGCCG	GTGCGTATT	CGGACTATCC
	151	GTCCCCGATC	CGCGCCCTGT	TGGTGCAGAA	GGACGTGCCG	ATGACGGTTT
	201	TGTTTGTGTT	GTCAGCGTGC	TCGATGCTGA	CGGTTTGATC	AAAATCCTGA
15	251	CGCATACCGT	TAAAAATTTT	ACCGGTTTCG	CGCCGTTGGG	AACGGTGTTC
	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC
	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAACTC	ACTACTTTTA
	401	TGGTTGTTTT	TACAGGGATT	TTATCTAATA	CCGCTTCTGA	ATTGGGCTAT
	451	GTCGTCTTAA	TCCCTTTGTC	CGCCATCATC	TTTCATTCCC	TCGGCCGCCA
20	501	TCCGCTTGCC	GGTCTGGCTG	CGGCTTTCGC	CGGCGTTTCG	GGCGGTTATT
	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTTGCC	AGGCATCACC
	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTATC	TGAAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
25	751	GATTTGTGAC	AAGAAGAAAA	AGACATTCGA	CATTCCAATG	AAATCACGCC
	801	TTTGGATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTTC	GTTGCCTTAT
	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCCTCAT
	901	CCTGAAACAG	GATTGGTTTC	CGGTTCCGCG	TTTTTAAAT	CAATTGTTGT
	951	TTTTATTTC	TGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
30	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG
	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTTCG	CCGCACAGTT
	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	CGCGTTAAAG
	1151	GGCGGACGTT	CTTAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTTATCAAT	CTGATGATAG	GCTCCGCCTC
35	1251	CGCGCAATGG	GCGGTAATCG	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGG	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTTCTA
	1451	TGATGTTGCC	GTATTCGCGT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
40	1501	TGCATTTGGG	TATTTGTTT	GGGCCTGCCC	GTCGGTCCCG	GCGCGCCAC
	1551	ATTCTATCCC	GCACCTTAA			

This encodes a protein having amino acid sequence <SEQ ID 138>:

	1	MSQTDTRQDGR	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS
	51	VDPDRPVGAK	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL
45	101	VSLLGVGIAE	KSGLISALMR	LLLTSPRKL	TFMVFVFTGI	LSNTASELGY
	151	VVLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANLEFG	TIDPLLAGIT
	201	QQAQIITHPD	YVVGPEANWF	FMVASTFVIA	LIGYFVTEKI	VEPQLGPYQS
	251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
	301	PETGLVSGSP	FLKSIVVFIF	LLFALPGIVY	GRVTRSLRGE	QEVVNMAES
50	351	MSTLGLYLV	IFFAAQFVAF	FNWTNIGQYI	AVKGATFLKE	VGLGGSVLEI
	401	GFILICAFIN	LMIGSASQW	AVTAPIFVPM	LMLAGYAPEV	IQAYRIGDS
	451	VTNIITPMMS	YFGLIMATVI	KYKDGAVGT	LISMMLPYSA	FFLIAWIALF
	501	CIWVFLGLP	VGPGAPTFYP	AP*		

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

		10	20	30	40	50	60
	orf12a.pep	MSQTDTRQDGR	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS	VDPDRPVGAK
60	orf12-1	MSQTDTRQDGR	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAVAGAYFGLS	VDPDRPVGAK
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf12a.pep	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL	VSLLGVGIAE	KSGLISALMR
65	orf12-1	GRADDGLIY	IVSLNADGFI	KILTHTVKNF	TGFAPLGTVL	VSLLGVGIAE	KSGLISALMR

		70	80	90	100	110	120
5	orf12a.pep	130	140	150	160	170	180
		LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS					
	orf12-1	130	140	150	160	170	180
		LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS					
10	orf12a.pep	190	200	210	220	230	240
		GGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
	orf12-1	190	200	210	220	230	240
		GGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
15	orf12a.pep	250	260	270	280	290	300
		VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
	orf12-1	250	260	270	280	290	300
		VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH					
20	orf12a.pep	310	320	330	340	350	360
		PETGLVSGSPFLKSIVVFIFFLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLV					
	orf12-1	310	320	330	340	350	360
		PETGLVSGSPFLKSIVVFIFFLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLV					
25	orf12a.pep	370	380	390	400	410	420
		IFFFAAQFVAFENWNTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW					
	orf12-1	370	380	390	400	410	420
		IFFFAAQFVAFENWNTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQW					
30	orf12a.pep	430	440	450	460	470	480
		AVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
	orf12-1	430	440	450	460	470	480
		AVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT					
35	orf12a.pep	490	500	510	520		
		LISMMLPYSAFFLIAWIALFCIWVFLVGLPVGPGAPTFFYPAPX					
	orf12-1	490	500	510	520		
		LISMMLPYSAFFLIAWIALFCIWVFLVGLPVGPGAPTFFYPAPX					

gonorrhoeae:

		orf12.pep	orf12.ng	
50	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI	30	
	orf12.ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMAASTFVIALI	232	
55	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90	
	orf12.ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292	
60	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150	
	orf12.ng	PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMS	352	
65	orf12.pep	TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210	
	orf12.ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAFLKKFRLGGSVLFIFIGFILICAFINLM	412	
65	orf12.pep	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAAYRIGDSVTNIITPMSYFGLIMATVXXY	270	
	orf12.ng	IGSASAQWAVTAPIFVPMMLLAGNAPQVQAAAYRIGDSVTNIITPMSYFGLIMATVIKY	472	

```

orf12.pep      KKDAGVGTLLIXMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAP  320
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng        KKDAGVGTLLISMLPYSAFFLIWIALFCIWVFLGLPVGPGTPTFFYPVP  522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
      51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTT ATTATTTTCA
    101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATTT CGGACTATCC
    151  GTCCCCGATC CGCGTCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
    201  GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
    251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTGGG AACGGTGTG
    301  GTTTCCTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGC
    351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAATC ACTACTTTTA
    401  TGGTTGTTTT TACAGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
    451  GTCGTCTCTAA TCCCTTTGTC CGCCGTCTAT TTTTCATCGC TCGGCCGCCA
    501  TCCGCTTGCC GGTTTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
    551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
    601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
    651  CAAGTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
    701  ATTTTGTAC TGAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
    751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
    801  TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTG GTGCCTTAT
    851  CCGCCCTATT GGCTTGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
    901  CCTGAAACAG GATTGGTTGC CGGTCGCGG TTTTAAAT CGATTGTGTG
    951  TTTTATTTTC TTGTTGTTG CGCTGCCGGG CATTGTTTAT GGCCGGATAA
    1001  CCCGAAGTTT GCGCGCGGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
    1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
    1101  TGTCGCATTT TTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
    1151  GGGCGGTGTT CTTAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTTTATC
    1201  GGTTTTATTT TAATTGTGTC TTTTATCAAT CTGATGATAG GCTCCGCCTC
    1251  CGCGCAATGG GCGGTAATG CGCCGATTTT CGTCCCTATG CTGATGTTGG
    1301  CCGGTACGCG GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
    1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTCGGGC TGATTATGGC
    1401  GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCAGC CTGATTCTA
    1451  TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
    1501  TGCATTGGG TATTTGTTT GGGTCTGCC GTCGGTCCC GCACACCAC
    1551  ATTCTATCCG GTGCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

      1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
    51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
    101  VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LNSTASELGY
    151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
    201  QQAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
    251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
    301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
    351  MSTLGLYLVI IFFAAQFVAF FWNWNIGQYI AVKGAVFLKK FRLGGSVLEI
    401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
    451  VTNIIPTMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
    501  CIWVFLGLP VPGTPTFFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orf12-1.pep  MSQTDQTDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
orf12ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
      10      20      30      40      50      60
55      70      80      90     100     110     120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR
orf12ng      GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSLGISALMR
      70      80      90     100     110     120
60
      130     140     150     160     170     180
orf12-1.pep  LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIFHSLGRHPLAGLAAAFAGVS
orf12ng      LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIVFHSLSGRHPLAGLAAAFAGVS
65

```

-133-

		130	140	150	160	170	180
5	orf12-1.pep	190	200	210	220	230	240
	orf12ng	190	200	210	220	230	240
10	orf12-1.pep	250	260	270	280	290	300
	orf12ng	250	260	270	280	290	300
15	orf12-1.pep	310	320	330	340	350	360
	orf12ng	310	320	330	340	350	360
20	orf12-1.pep	370	380	390	400	410	420
	orf12ng	370	380	390	400	410	420
25	orf12-1.pep	430	440	450	460	470	480
	orf12ng	430	440	450	460	470	480
30	orf12-1.pep	490	500	510	520		
	orf12ng	490	500	510	520		

In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

40	sp P46133 YDAH ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION >gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli] Length = 510 Score = 329 bits (835), Expect = 2e-89 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)
45	Query: 8 RSGRFLRTVEWLG NMLPHPV TXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67 +SG+ VE +GN +PHP +A+ +FG+S +P D Sbjct: 13 QSGKLYGWVERIGNKVPHPFLLEFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68 IHVVSLLDADGLIKILTHTVKNETGFAPXXXXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127 + V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + + Sbjct: 65 VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124
55	Query: 128 RKLTTFMVVF TGI LSNTASELG YVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187 + ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL Sbjct: 125 ARYASYMVLFI AFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVGC GFTANL 184
60	Query: 188 FLGTIDP LLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247 + T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG Sbjct: 185 LIVTTDVL LSGISTEAAA AFNPQMHSVIDNWYFMASVVVLTIVGGLITDKIIEPLGQ 244
65	Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALLAWSIVPADGILRHPE TGLVA 307 +Q + ++ + + S GL AGVV + A +A ++P +GILR P V Sbjct: 245 WQGSDEKLTQLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298
70	Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXXXX 367 SPF+K IV I L F + + YG TR++R + ++ + M E M + ++ Sbjct: 299 PSPFIKIVPLIILFFVVS LAYGIATR TIRRQADLPHLMIEPMKEMAGFIVMVFP LAQF 358
	Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASQAQWAVTAPIF 427 NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

-134-

Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSIAIWSILAPIF 418

Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLSMMMLP 487

VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP

Sbjct: 419 VPMFMLLGfHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDakLGTYYSLVLP 478

Query: 488 YSAFFLIAWIALFCIWVFLVGLPVGPG 514

Y FL+ W+ + W +++GLP+GPG

Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```

1  ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTTCGTAA CGGACAGTCA
51  GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAACA GGTTTTTTTC
101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151 ATgGCTTCGC GCAGTGCCTC TATACCGTA TTTTCAGCAA CGGAAATGCG
201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTgTT
251 CTTCAGACGG CAGCAGGTCG GTTTGTGTGT ACACCTTgAT GCACGGAAaTA
301 TCGCCGGCAT GGATTCTTGT CAGTACGTTT TCCACGTCTT CAATCTGCTG
351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
25  401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTCCGGC
451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTGGGG
501 ACT..

```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```

1  ..TAGAAGXXVF VFVTDsQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA
51  MASRSASIPV FSATeMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI
101 SPAWISCSTF STSSICPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
151 RXLTNPtVSv RIMLHSG..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

				10	20	30
orfl4.pep				TAGAAGXXVFVFVTDsQVEVFGNIQTAVET		
				: : : : : :		
40 orfl4a	GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFVTDGQMqVFGNVQPAVET	150	160	170	180	200
		40	50	60	70	80
45 orfl4.pep	GFFHGISVSSVFGAAQDSAMASRSASIPVFSATeMRTAAIFPAASRHMPVFCSSDGSRS					
orfl4a	GFFHGISVSSVFGAAQYSAMASRSASIPVFSATeMRTAAIFPAASRHMPVFCSSDGSRS	210	220	230	240	250
50 orfl4.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG	100	110	120	130	140
orfl4a	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG	270	280	290	300	310

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAGG
	51	TCGGCAGCGC	GAACATCATC	GTCTGCATCA	TCCCCAGCCC	GGCAACGCGC
10	101	AGGCGGACGA	TGTATTGTTT	CGCTTCTTTT	TGGTGGCGG	CTTCGATTTT
	151	TTGCGCGTCA	TGGGTGCGG	CGGTGTAGCC	TATGTCCTG	ATTTTCAACA
	201	GAATGTCGGA	AAGGCGGATT	TTGCCGTCGT	CCCAGACGAC	GCGGCAGCGG
	251	TGCGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCGT	TACGCAAAAG
	301	CTGCTGTTCT	ATCAGCCAGA	CGCAGGCGCG	GCCAGTGATG	CCGCGGAGCA
15	351	TTAAAACCGC	CTCGCGGTG	CGCCCGTGGG	TTTCCACAAA	GTCGGAGTCG
	401	ACTTCGGGCA	GGTCGTACAG	GCGGATTGCG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTTGTAA	TAACTGCCCC
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTCGCGGTT	CTTCGTTTTC	GTAACGGACG	GTCAGATGCA	GGTTTTCGGG
20	601	AACGTCCAGC	CCGCAGTGGA	AACAGGTTTT	TTTCATGGCA	TTTCGGTTTT
	651	GTCTGTGTTT	GGTGCGCGCG	CACAATACTC	GGCAATGGGT	TCGCGCAGTG
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GCGCAATTTT
	751	CCGCGAGCGT	CGCGCCATAT	GCCCGTGTTT	TGTTCTTCAG	ACGGCAGCAG
	801	GTCGGTTTTG	TTGTACACCT	TGATGCACGG	AATATCGCCG	GCATGGATTT
25	851	CTTGCACTAG	GTTTTCCACG	TCTTCAATCT	GCTGTCCGCT	GTTCCGAGCG
	901	GCGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC
	1001	CGACGCGTAT	GGTCAGGATA	ATGCTGCATT	CGGGACTTGT	GTACAGCCGC
	1051	CGCGCCGTCG	TGTCGAGTGT	GGCGAAAAGC	TGGTCTTTGC	CATATATGCC
	1101	CGACTTGGTC	AGCCGTTTGA	ACAGACTGGA	TTTGGCGACA	TGGGTATAG

	1	MEDLQEI GFD	VAAVKVGRQR	EHHRLHHPQP	GNGEADDVLF	AFFLVGGFDF
	51	LRVIGCGGVA	YLPDFQONVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
	101	LLFDQPDAG	AGDAAEH*NR	LARAAVG FHK	VGLDFGQVVQ	ADLVDFELGR
35	151	LGFLRLVGG	LVITAQAARV	NNALCDLTT	GAAGFVVFV	VDGGMQVFG
	201	NVQPAVETGF	FHGISVSSVF	GAAQYSAMA	SRSASIPVFS	ATEMRATAIF
	251	PAASRHMPVF	CSSDGRSVL	LYTLMHGISP	AWISCTSTFT	SSICCP LFGA
	301	AASTTCSSTS	ACAVSSSVAE	KAEISLCGRS	LTNPTVSVRI	MLHSGLMYSR
	351	RAVSSVAKS	WSFAYMPDLV	SRLNRLD LPT	LV*	

40 Homology with a predicted ORF from *N.gonorrhoeae*

	orf14.pep	TAGAAGXXVFVFTDSQVEVFGNIQTAVET	30
		: : : : : : : : :	
45	orf14ng	GRQGFRRVGGASFVITAQAGIDDALCCLTADAAGFAVFAFVADGQMQVFGNVQPAVET	208
	orf14.pep	GFHHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSR	90
50	orf14ng	GFHHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSR	268
	orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAIEISLCG	150
	orf14ng	VLLYTLMHGISWAWISCSTFSTSSICPLFRAAASTTCSSTSACTVSSKVAEKAIEISLCG	328
55	orf14.pep	RXLTNPTVSVRIMLHSG	167
	orf14ng	RSLTNPTVSVRIMLHAGLMYSRRVAVSRVAKSWSFAYMPDLVSRNLRLDLPTLV	382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
 51 LRVIGCGGVA CLPDFQONVG EADFAVVPDD AAARAVIEV DADDAVCAQK
 101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
 151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMVFVG
 5 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
 201 NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
 251 PAASRHPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLEFRA
 301 AASTCSSTS ACTVSSKVAE KAEISLCGRS LTNPVTSVRI MLHAGLMYSR
 351 RAVSVRAKS WSFAYMPDLV SRLNRLDLPT. LV*

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the
 10 proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for
 vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

15 1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
 51 GCGGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA
 101 TGCCGAATC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GCGCGCTTTG
 151 TCGTTCGGCG CGCTGATGAT TGCCTGTGTA GACGTGTCGT CAAATATGGC
 201 GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
 251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
 20 GTGGCGGCGA TTTGCGCGTT TGTGTTTGG TATATCGGTT TGGCGAACAC
 351 CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTGTTGGCG TTTTATGTGG
 401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
 451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCGCG
 501 GAATCAGGAA AAAGCCAAC TGGATCGCACT CTTAAAA.CC GCGC..

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

1 ..GHYSDRTWKP RLXGRRLLPYL LYGTIAVIV MILMPNSGSF GFGYASLAAL
 51 SFGALMIALL DVSSNMAMQP FKMMVGDMVN EEQKXYAYGI QSFLANTGAV
 101 VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
 151 EYXPETYARY HGIDVAANQE KANWIALLKX A..

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

1 ATGTCGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
 51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
 101 CCTTTACCCCT GCAAAGCTCG CAAATGAGCC GCATTTTTC AACGCTAGGC
 151 GCAGACCCGC ACAATTGGG CTGGTTTTC ATCCTGCCGC CGCTGGCGGG
 35 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
 251 CGCGTTTGGG CGGCCGCGCT CTGCCGTATC TGCTTTATGG CACGCTGATT
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
 351 CTATGCGTCG CTGGCGGCTT TGTCGTTCGG CGCGCTGATG ATTGCGCTGT
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
 40 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
 501 CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
 601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
 651 GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAACCC TACGCCGTT
 45 701 ACCACGGCAT CGATGTCGCC GCGAATCAGC AAAAAGCCAA CTGGATCGAA
 751 CTCTTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT
 801 CTTCTGCTGG TTCGCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
 851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTCCGT AGGTTATCAG
 901 GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCACT CGGTTGCGGC
 50 951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCGGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CCTTGTGCGG CAAGCATATG GGCACCTACT TGGGCTTGTT TAACGGCTCT
 55 1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC
 1301 TGCTGCTGGG CGCGTTTTCG GTGTTCTGTA TTAAAGAAAC ACACGGCGGG
 1351 GTTTGA

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

1  MSEYTPQTAK QGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG
51  ADPHNLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151 DMVNEEQKY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201 VVVAFYVGAA LLVITSFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYYS AGAIAENVWH TTDASSVGYQ
301 EAGNWSYGLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLENFS
10 401 ICMPOIVASL LSFVLFPLMG GLQATMFLVG GVVLLGAFS VFLIKETHGG
451 V*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
orfl6.pep                        GHYSDRTWKPRXLGRRLPYLLYGTLIIVIV
20 orfl6a                        IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGTLIIVIV
                                     50      60      70      80      90     100

                                     40      50      60      70      80      90
orfl6.pep                        MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKYAYGI
25 orfl6a                        MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKYAYGI
                                     110     120     130     140     150     160

                                     100     110     120     130     140     150
orfl6.pep                        QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAAALLVITSFTIIFVK
30 orfl6a                        QSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAAALLVITSFTIIFVK
                                     170     180     190     200     210     220

                                     160     170     180
orfl6.pep                        EYXPETYARYHGIDVAANQEKANWIALKXA
35 orfl6a                        EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYYSAGAI
                                     230     240     250     260     270     280

40 orfl6a                        AENVVHTTDASSVGYQAEAGNWSYGLAAVQSVAVICSFVLAKVPNKYHKAGYFGCLALGA
                                     290     300     310     320     330     340

```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
45 51  AAAAGACAGC ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101 CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT
151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
251 CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301 CGCGTTATTG TGATGATTTT GATGCCGAAC TCGGCGAGCT TCGGTTTCGG
50 351 CTATGCGTCG CTGGCGGCTT TGTGTTTCGG CGCGCTGATG ATTGCGCTGT
401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501 CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGAGACC
601 GTGGTCGTGG CGTTTATGT GGGTCCGGCG TTGCTGGTGA TTACCAGCGC
651 GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCCGTT
701 ACCAGGCGAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751 CTCTTGAAAA CCGCGCCTAA GGCGTTTGGG ACGGTTACTT TGGTGCAATT
801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
851 TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
901 GAGGCGGGTA ACTGGTACGG CGTTTGGCGG GCGGTGCAGT CGGTTGCGGC
951 GGTGATTGTG TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG

```

-138-

1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATTG TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CCTTGTCTGG CAAGCATATG GGCCTTACT TGGGCTGTT TAACGGCTCT
 5 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCCTCGTCC
 1301 TGCTGCTGGG CGCGTTTTCG GTGTTCTCTGA TTAAGAAAC ACACGGCGGG
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10 1 MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
 51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGGRR LPYLLYGTLI
 101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
 151 DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
 201 VVVAFYVGAA LLVITSAFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
 15 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVVH TTDASSVG YQ
 301 EAGNWyGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
 351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS
 401 ICMPIQIVASL LSFVLFPM LG LOATMFLV GVVLLGAFS VFLIKETHGG
 451 V*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
	orf16a.pep	MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF					
	orf16-1						
25		10	20	30	40	50	60
	orf16a.pep	ILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGTLI					
	orf16-1						
30		70	80	90	100	110	120
	orf16a.pep	ILPPLAGMLVQPIVGHYSDRTWKPRLGRRLPYLLYGTLI					
	orf16-1						
35		130	140	150	160	170	180
	orf16a.pep	LAALSFGALMIALLDVSSNMAMQPFKMMVGD					
	orf16-1						
40		190	200	210	220	230	240
	orf16a.pep	FVFAYIGLANTA					
	orf16-1						
45		250	260	270	280	290	300
	orf16a.pep	ANQEKANWIELLKTAPKAFWTVTLVQFFCW					
	orf16-1						
50		310	320	330	340	350	360
	orf16a.pep	EAGNWyGVLA					
	orf16-1						
55		370	380	390	400	410	420
	orf16a.pep	LSYTLIGIAWAGIITYPLTI					
	orf16-1						
60		370	380	390	400	410	420
	orf16a.pep	GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
	orf16-1						
65		430	440	450			
	orf16a.pep	GLQATMFLVGGVVLLGAFSVFLIKETHGGVX					
	orf16-1						
		430	440	450			

5	orf16.pep	GHYSDRTWKPRXLGRRLPYLLYGTLLIAIVV	30
		:	
	orf16ng	HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLGRRLPYLLYGTLLIAIVV	131
10	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQFFKMMVGMVNVEEQKXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQFFKMMVGMVNVEEQKSYAYGI	191
15	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK	150
	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLIITSAFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVTPVQFFCWFAFRYMWYTSAGAI	311

	1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
	51	TACTTTTCAA	ATCAAAAAAA	AGGATTTACT	TTATGTCGGA	ATATACGCCT
	101	CAAACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTTGGAT
25	151	GTTGAGCTTC	GGCTATCTCG	CGGTTACAGC	GGCCTTTACC	CTGCAAAAGCT
	201	CGCAGATGAG	CCGCATTTTT	CAAAACGCTAG	GGCGAGACCC	GCACAATTTG
	251	GGCTGGTTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTCAGCCGAT
	301	AGTGGCTACT	ACTCAGACCG	CACTTGGAAG	CCGCGCTTGG	GCGGCCGCCG
	351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTATATC	GTGATGATTT
30	401	TGATGCCGAA	CTCGGGCAGC	TTGCGTTTCG	GCTATGCGTC	GCTGGCGGCC
	451	TTGTCGTTTC	GCGCGCTGAT	GATTGCGCTG	TTGGACGTGT	CGTCGAATAT
	501	GGCGATGCAG	CCGTTTTAAGA	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
	551	AGAAAAGCTA	CGCTTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACCGG
	601	GTTGTGGCAG	GCAATCTGCC	GTTTGTGTTT	GGGTATATCG	TTTGGCGGAA
35	651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA	GCATTCTATG
	701	TGGGTGCGCG	GTTACTGATT	ATTACCACTG	CGTTCACAA	CTCCAAAGTC
	751	AAAGAAATAC	ACCCGGAAC	CTACGCCCGT	TACCACGGCA	CCGATGCTGC
	801	CGCGAATCAG	GAAAAAGCCA	ACTGGTTCGA	ACTCTTAAAA	ACCGCGCTCA
	851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG	GTCGCGCTTC
40	901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGTCTGGCA
	951	CCTACCAGAT	GCGTCTTCCG	TAGGCCATCA	GGAGCGGGGC	AACCGGTACG
	1001	GCGTTTTGGC	GGCGGTGTAG			

	1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLLLYVG	IYASNSKTRF	ARAGKKHDL
	51	VELRLSRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHPA	AAGGDAGSAD
45	101	SGYVSDRTW	PRLGRRRLPY	LLYGTLIAVI	NMILMPNSGS	FGFGYASLAA
	151	LSFGALMIAL	DLVSSNMAMQ	PFKMMVGDV	VEEQKSAYYG	IQSFLANTDA
	201	VVAAILPEVF	AYIGLANTAE	KGVVVPQTVV	AEYVGAALLI	ITSAFTISKV
	251	KEYDPETYAR	YHGDVVAANQ	EKANWFELLK	TAPKVFEVTV	PVQFFCWFAP
	301	RYMWTYSAGA	IAENGVHHTD	ASSVGHQEAG	NRKGVLAAV*	

```

55      30      40      50      60      70      80
orf16-1.pep  MLSEFGLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSVRT
                        |  ::  |  ||  :  |::|||
orf16ng      DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGDGASDSDGYYSVRT
50      60      70      80      90      100

      90      100      110      120      130      140
orf16-1.pep  WKPRLGRRRLPYLLYGTLIAVIVMILMPNSGSGFGGYASLAALSFGALMIALLDVSSNMA
                |||||
60      orf16ng  WKPRLGRRRLPYLLYGTLIAVIVMILMPNSGSGFGGYASLAALSFGALMIALLDVSSNMA
                |||||
110      120      130      140      150      160

```

-140-

		150	160	170	180	190	200
5	orf16-1.pep	MQPFKMMVGDVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEGVVPQTV					
	orf16ng	MQPFKMMVGDVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTV					
		170	180	190	200	210	220
10	orf16-1.pep	VVAFYVGAALLVITSFTIFKVKEYDPETYARYHGIDVAANQEKANWIELLKTAPKAFWT					
	orf16ng	VVAFYVGAALLIITSFTISKVKEYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWT					
		230	240	250	260	270	280
15	orf16-1.pep	VTLVQFFCFWFAFYQYMWYTSAGAIAENVWHTTDASSVGYQEAGNRYGVLAQVSVAAVICS					
	orf16ng	VTPVQFFCFWFAFRYMWYTSAGAIAENVWHTTDASSVGHQEAGNRYGVLAQVX					
		290	300	310	320	330	340

- 20 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAGG	GGCAGCCTGG	TGATGATGGG
30	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AA.NTGACGG
	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCTT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTG	GATCGNCTGG
	351	CAGCCAGAAT...				

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNPN	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVEDSA	XXTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN...			

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

40	1	ATGTTGTTCC	GTAAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAGG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCTT	TCCAAATAGT	TGAGGATACC
45	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTG	AATCGCCTGG
	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCGA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAAATA
	501	CTACGCCACA	CCGCAAAAAC	TGAACGCCGA	TTACCATTTT	GAGCAAAGTG
50	551	TGCCTGCCGA	TATTTATTAC	ACGGTTACTG	AAGAACATAC	CGACAAATCC
	601	AAGCTGTTTG	CAATATATCTT	ATATACGCCC	CCCTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTACTGGCCT	TGCCTGCGGC	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAATGA			

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

55	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNPN	VSETITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WVFNVEDSA	KLTGILKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKLESPGSON	FSTEGLCRLR	DTDKPADIAT	LKQLGFCAVK
	151	LDNRITYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVPADIYY	TVTEEHTDKS

201 KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N.*

5 *meningitidis*:

		10	20	30	40	50	60
orf28.pep		MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK					
		: : : : : :					
10 orf28a		MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf28.pep		GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGXSQSON					
		: : : : : :					
15 orf28a		GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN					
		70	80	90	100	110	
orf28a		FSTEGLCRLRYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		120	130	140	150	160	170

20 The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGTT
	51	GAACGCTGT	ACGGTAATGA	TGTGGGGTAT	GAACAGCCCG	TTACAGCGAAA
	101	CGACCGCCCG	CAAACACGTT	GACAAGGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTGCGCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
25	201	CGGGAAATAC	TGGTTCGTGC	TCAATCCTGA	AGATTCCGGC	AAGCTGACGG
	251	GCATTTTGAA	GGCCGGGTTG	GACAAGCAGT	TTCAAATGGT	TGAGCCCAAC
	301	CCGCGCTTTG	CCTACCAAGC	CCTGCCGCTC	AAACTCGAAT	CGCCCGCCAG
	351	CCAGAATTTT	AGTACCGAAG	GCCTTTGCCT	GCGCTACGAT	ACCGACAGAC
	401	CTGCCGACAT	CGCCAAGCTG	AAACAGCTTG	AGTTTGAAGC	GGTCGAACTC
30	451	GACAATCGGA	CCATTACAC	GCGCTGCGTC	TCCGCCAAAG	GCAAATACTA
	501	CGCCACACCG	CAAAAACCTGA	ACGCCGATTA	TCATTTTGAG	CAAAGTGTGC
	551	CTGCCGATAT	TTATTACACG	GTTACGAAAA	AACATACCGA	CAAATCCAAG
	601	TTGTTTGAAA	ATATTGCATA	TACGCCACC	ACGTTGATAC	TGGATGCGGT
	651	GGGCGCGGTG	CTGGCCTTGC	CTGTCGCGGC	GTTGATTGCA	GCCACGAATT
35	701	CCTCAGACAA	ATGA			

This encodes a protein having amino acid sequence <SEQ ID 160>:

	1	MLFRKTTAAV	LAATLMLNGC	TVMWGMNSP	FSETTARKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFVNPEDSA	KLTGILKAGL	DKQFQMVPEPN
	101	PRFAYQALPV	KLESPASQNF	STEGLCRLRYD	TDRPADIAKL	KQLEFEAVEL
40	151	DNRTIYTRCV	SAKGKYYATP	QKLNADYHFE	QSVADIYTT	VTKKHTDKSK
	201	LFENIAYTPT	TLILDVAGAV	LALPVAALIA	ATNSSDK*	

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

		10	20	30	40	50	60
45 orf28a.pep		MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK					
		: : : : : :					
orf28-1		MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
		10	20	30	40	50	60
		70	80	90	100	110	119
50 orf28a.pep		GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN					
		: : : : : :					
orf28-1		GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPASQN					
		70	80	90	100	110	120
55 orf28a.pep		120	130	140	150	160	170
		FSTEGLCRLRYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		: : : : : :					
orf28-1		FSTEGLCRLRYDTRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		130	140	150	160	170	180

-142-

```

180      190      200      210      220      230
orf28a.pep  EQSVPADIYYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX
5 orf28-1    EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
              190      200      210      220      230

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*

10 *gonorrhoeae*:

```

orf28.pep  MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVAEDNAQLEK 60
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
orf28ng    MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSTITRKHVDKDQIRAFGVVAEDNAQLEK 60

15 orf28.pep  GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGXGSGQN 120
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
orf28ng    GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSGQN 120

```

The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

```

20      1  ATGTTGTTC  GTAAACGAC  CGCCGCCGT  TTGGCGGCA  CCTGATACT
51  GAACGGCTGT  ACGATGATGT  TCGGGGGGAT  GAACAACCG  GTCAGCCAAA
101 CAATCACCCG  CAAACACGTT  GACAAAGACC  AAATCCGCG  CTTCGGTGTG
151 GTTGCCGAAG  ACAATGCCCA  ATTGGAAGAG  GGCAGCCTGG  TGATGATGGG
201 CGGGAAATAC  TGGTTCGCCG  TCAATCCCGA  AGATTCGGCG  AAGCTGACGG
251 GCCTTTTGAA  GGCCGGGTTG  GACAAGCCCT  TCCAAATAGT  TGAGGATACC
25  301 CCGAGCTATG  CCGGCCACCA  AGCCCTGCCG  GTCAAATCG  AAGCGCCCGG
351 CAGCCAGAAT  TTCAGTACCG  GAGGTCTTTG  CCTGCGCTAT  GATACCGGCA
401 GACCTGACGA  CATCGCCAAG  CTGAAACAGC  TTGAGTTTAA  AGCGGTCAAA
451 CTCGACAATC  GGACCATTTA  CACGCGCTGC  GTATCCGCCA  AAGGCAAATA
501 CTACGCCACG  CCGCAAAAC  TGAACGCCGA  TTATCATTTT  GAGCAAAGTG
30  551 TGCCCGCCGA  TATTTATAT  ACGGTTACTG  AAAACATAC  CGACAAATCC
601 AAGCTGTTTG  GAAATATCTT  ATATACGCCC  CCCTTGTTGA  TATTGGATGC
651 GGCGGCCGCG  GTGCTGGTCT  TGCCTATGGC  TCTGATTGCA  GCCGCGAATT
701 CCTCAGACAA  ATGA

```

This encodes a protein having amino acid sequence <SEQ ID 162>:

```

35      1  MLFRKTTAAV  LAATLILNGC  TMMLRGMNPN  VSQTITRKHV  DKDQIRAFGV
51  VAEDNAQLEK  GSLVMMGGKY  WFAVNPEDSA  KLTGLLKAGL  DKPFQIVEDT
101 PSYARHQALP  VKFEAPGSGN  FSTGGLCLRY  DTGRPDIAK  LKQLEFKAVK
151 LDNRTIYTRC  VSAKGKYAT  POKLNADYHF  EQSVPADIYY  TVTEKHTDKS
201 KLFGNILYTP  PLLILDAAA  VLVLPMALIA  AANSSDK*

```

40 ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

```

          10      20      30      40      50      60
orf28-1.pep  MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
45 orf28ng    MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSTITRKHVDKDQIRAFGVVAEDNAQLEK
          10      20      30      40      50      60

          70      80      90      100     110     120
orf28-1.pep  GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSGQN
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
50 orf28ng    GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSGQN
          70      80      90      100     110     120

          130     140     150     160     170     180
orf28-1.pep  FSTGGLCLRYDTDKPADIAKLKQLGFCAVKLDNRTIYTRCVSAKGKYATPQKLNADYHF
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
55 orf28ng    FSTGGLCLRYDTGRPDIAKQLEFKAVKLDNRTIYTRCVSAKGKYATPQKLNADYHF
          130     140     150     160     170     180

          190     200     210     220     230     239
60 orf28-1.pep  EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
          |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:| |||:|||||:|
orf28ng    EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAALVLPMALIAAANSSDKX

```

190 200 210 220 230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```

15      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
      101 CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
      151 GACGGCGGTT TTAAGTGTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
      201 TCCGAGGAGT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCGG
      251 GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
      301 ACAAGAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGCTAGTA
      351 AGAAATGCC GGTGCCGCCT CTGGT..

```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```

25      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHDS KSTSDFSGGV
      51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
      101 TKTSIVPQAP FSDRWLEENA GAASG..

```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```

30      1  ATGAATTGTC CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC
      51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
      101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
      151 TTTGGTAATG CTCGCGCGAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
      201 ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
      251 CAGGGTTTGA AGGTGTATAT GGTATGAAA CCCATTTTTC AGGGCACGGA
      301 CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
      351 TTTCAGCGGC GGTGTAGACG GCGGTTTTAC TGTTTACCAA CTTTCATCGAA
      401 CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
      451 GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
      501 AGGAACCTCA ACAAAAACAA AGACTAATAT TGTCCTCAA GCCCATTTT
      551 CAGACCGTTG GCTAAAAGAA AATGCCGTG CCGCCTCTGG TTTTTCAGC
      601 CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
      651 TTGGTGGGCT AACCGTATGG ATGATGTTTC CGGCATCGTC CAAGGTGCGG
      701 TTAATCCTTT TTTAATGGGT TTCAAGGAG TAGGGATTGG GGCAATTACA
      751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
      801 AGGTATTAAAT GATTTAGGAA AATTAAGTCC GGAAGCACA CTTGCTGCCG
      851 CGAGCCTATT ACAGGACAGT GCTTTTGGCG TAAAAGACGG TATCAACTCT
      901 GCCAAACAAT GGGCTGATGC CCATCCAAT ATAACAGCTA CTGCCCAAAC
      951 TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTCGAGA GGTAAAAAAG
      1001 TAGAACTTAA CCGACTAAA TGGGATTGGG TAAAAATAC CGGTTATAAA
      1051 AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
      1101 GAATAAACCT ATTAAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC
      1151 AAAATTTTGA GAAGTTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA

```


10

15

Computer analysis of this amino acid sequence gave the following results:

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

	1	ATGAATTNGC	CTATTCAAAA	ATTCATGATG	CTGTTTGCAG	CAGCAATATC
45	51	GTNGCTGCAA	ATCCCNATTA	GTCATGCGAA	CGGTTTGGAT	GCCCGTTTGC
	101	GCGATGATAT	GCAGGCCAAA	CACTACGAAC	CGGGTGGTAA	ATACCATCTG
	151	TTTGGTAAATG	CTCGCGCGAG	TGTTAAAAAT	CGGGTTTACG	CCGTCCAAAC
	201	ATTTGATGCA	ACTGCGGTCTG	GCCCCATACT	GCCTATTACA	CACGAACGGA
50	251	CAGGATTTGA	AGGCATTATC	GGTTATGAAA	CCCATTTTTC	AGGCACATGGA
	301	CATGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAA	GCACCTCTGA
	351	TTTCAGCGGC	GCGGTAGACG	GTGGTTTTAC	CGTTTACCAA	CTTCATCGGA
	401	CAGGGTCGGA	AATCCATCCG	GAGGATGGAT	ATGACGGGCC	GCAAGGCAGC
55	451	GATTATCCG	CCCCCGGAGG	AGCAAGGGAT	ATATACANNT	ANATATGTCAA
	501	AGGAACITCA	ACAAAAACAA	AGAGTAATAT	TGTTCCCCGA	GCCCCATTAA
	551	CAGACCCCTG	CACTAAAGAA	AATGCCGCTG	CGCCCTCTGG	TTTTTTTCAGC
	601	CGTGCTGATG	AAGCAGGAAA	ACTGATATGG	GAAAGCGACC	CCAATAAAAA
60	651	TTGGTGGGCT	AACCGTATGG	ATGATATTTC	CGGCATCGTC	CAAGGTGCGG
	701	TTAATCCTTT	TTAATGGGT	TTTCAAGGAG	TAGGGATTGG	GGCAATTACA
	751	GACAGTGCAG	TAAGCCCGGT	CACAGATACA	GCCGCGCAGC	AGACTCTACA
	801	AGGTATNAAT	CATTTAGGAA	ANTTAAGTCC	CGAAGCACAA	CTTGCGGCTG
	851	CAACCGCATT	ACAAGACAGT	GCTTTTCGGG	TAAAAGACGG	TATCAATTCC
	901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAAGTCGAA	CAGCCCAAAA

-145-

5
10
15
20
25
30
35
40
45
50
55
60

```

951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGT TTGGGGC GGTA AAAAAG
1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TTA AAAATAC NGGCTATAAN
1051 ACACCTGCTG TTCGCACCAT GCATAC TTG GATGGGGAAA TGGCCGGTGG
1101 GAATAGACCG CCTAAATCTA TAACGTCCA CAGCAAAGCA GATGCTTCCA
1151 CACAACCGTC TTTACAAGCG CAACTAATTG GAGAACAAAT TANNNNNGG
1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTTAAA
1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAAT ATTGTTAGCC
1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT
1351 NATAAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG
1401 TACAGCATTT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTTATAG

```

This encodes a protein having amino acid sequence <SEQ ID 168>:

15
20
25
30
35
40
45

```

1 MNXPIQKFM LFAAAISXLQ IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS
201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT
251 DSAVSPVTD TAAQOTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGINS
301 ARQWADAHPN ITATAQTALA VAXAATTVWG GKKVELNPTK WDWVKN TGX
351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGE QIXXG
401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPXNMKE LPRGRTAYWD
451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL*

```

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

25
30
35
40
45
50
55
60

```

      10      20      30      40      50      60
orf29a.pep MNXPIQKFM LFAAAISXLQIPISHANGLDARLRDDMQAKHYEPPGGKYHLFGNARGSVKN
      10      20      30      40      50      60
orf29-1 MNLPIQKFM LFAAAISXLQIPISHANGLDARLRDDMQAKHYEPPGGKYHLFGNARGSVKK

      70      80      90     100     110     120
orf29a.pep RVYAVQTFDATAVGPILPITHERTGFEGIIIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG
      70      80      90     100     110     120
orf29-1 RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDHDSKSTSDFSG

      130     140     150     160     170     180
orf29a.pep GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGSTTKTKSNIVPR
      130     140     150     160     170     180
orf29-1 GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYSYYVKGSTTKTKTNIVPQ

      190     200     210     220     230     240
orf29a.pep APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG
      190     200     210     220     230     240
orf29-1 APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG

      250     260     270     280     290     300
orf29a.pep FQGVGIGAITDSAVSPVTD TAAQOTLQGXNHLGXLSPEAQ LAAATALQDS AFAVKDGINS
      250     260     270     280     290     300
orf29-1 FQGVGIGAITDSAVSPVTD TAAQOTLQGINDLGKLSPEAQ LAAASLLQDS AFAVKDGINS

      310     320     330     340     350     360
orf29a.pep ARQWADAHPNITATAQTALAVAXAATTVWGGKKVELNPTKWDWVKN TGXTPAVRTMHTL
      310     320     330     340     350     360
orf29-1 AKQWADAHPNITATAQTALSAEEAGTVWGGKKVELNPTKWDWVKN TGXKKPAARHMQTL

      370     380     390     400     410     420
orf29a.pep DGEMAGGNRPKPKSITSNSKADASTQPSLQAQLIGE QIXXGHAYNKHVIRQ QEFTDLNINS
      370     380     390     400     410
orf29-1 DGEMAGGNRPKPKSITSNSKADASTQPSLQAQLIGE QIXXGHAYNKHVIRQ QEFTDLNINS

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N. gonorrhoeae*:

5	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
10	orf29.pep	VHSPFDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDEYDGPQAAXYPPPGGARDIY	90
	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY	162
15	orf29.pep	SYVVKGTSTTKTTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTTKINTVPQAPFSDRWLEENAGAASGFLSRADEAGKLIWENDPKNWRANR	222

The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

1	MNLPIQKFMM	LFAAAISLLQ	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG
101	HEVHSPFDNH	DSKSTSDFSG	GVDGGFTVYQ	LHRTGSEIHP	EDGYDGPQGG
151	GYPPPGGARD	IYSYHIKGT	TKTKINTVPQ	APFSDRWLKE	NAGAASGFLS
201	RADEAGKLIW	ENDEDKNWR	NRMDDIRGIV	QGAVNPFITG	FQGLGVGAIT
251	DSAVSPVITYA	AARKTLQGIH	NLGNLSPEAQ	LAAATALQDS	AFAVKDSINS
301	ARQWADAHPN	ITATAQTALA	VTEAATTWVG	GKKVELNPAK	WDWVKNTGYK
351	KPAARHMQTV	DGEMAGGNKP	LESKNVTVTN	NFFENTGYTE	KVLRQASNGD
401	YHGFPPQSDA	FSENGTVIQI	VGGDNIVRHK	LYIPGSYKKG	DGNEFEYIREA
451	DGKINHRLFV	PNQQLPEK*			

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

1	atgAATTTGCG	CTATTCAAAA	ATTCATGATG	ctgttggcAg	cggcaatatac
51	gatgctGCat	ATCCCCATTA	GTCATGCGAA	CGGTTTGGAT	GCCCGTTTGC
101	GCGATGATAT	GCAGGCAAAA	CACTACGAAC	CGGGTGGCAA	ATACCATCTG
151	TTTGGTAATG	CTCGCGGCAG	TGTTAAAAAT	CGGGTTTTCG	CCGTCCAAAC
201	ATTTGATGCA	ACTGCGGTCG	GCCCCATACT	GCCTATTACA	CACGAACGGA
251	CAGGATTTGA	AGGTGTTATC	GGCTATGAAA	CCCATTTTTT	AGGACACGGA
301	CAGGAAGTAC	ACAGTCCGTT	CGATAATCAT	GATTCAAAAA	GCACCTTCTGA
351	TTTCAGCGGC	GGCGTAGACG	GCGGTTTTAC	CGTTTACCAA	CTTCATCGGA
401	CAGGGTCGGA	AATACATCCC	GCAGACGGAT	ATGACGGGCC	TCAAGGCGGC
451	GTTTATCCGG	AACCACAAGG	GGCAAGGGAT	ATATACAGCT	ACCATATCAA
501	AGGAACCTTA	ACCAAAACAA	AGATAAACAC	TGTTCCGCAA	GCCCTTTTTT
551	CAGACCGCTG	GCTAAAAGAA	AATGCCGGTG	CCGCTTCCGG	TTTTCTCAGC
601	CGTGCGGATG	AAGCAGGAAA	ACTGATATGG	GAAAACGACC	CCGATAAAAA
651	TTGGCGGGCT	AACCGTATGG	ATGATATTCG	CGGCATCGTC	CAAGGTGCGG
701	TTAATCCTTT	TTTAACGGGT	TTTCAAGGGG	TAGGGATTGG	GGCAATTACA
751	GACAGTGCGG	TAAGCCCGGT	CACAGATACA	GCCGCTCAGC	AGACTCTACA
801	AGGTATTAAAT	GATTTAGGAA	ATTAAAGTCC	GGAAGCACAA	CTTGCCGCCC
851	CGAGCCTATT	ACAGGACAGT	GCCTTTGCGG	TAAAAGACGG	CATCAATTCC
901	GCCAGACAAT	GGGCTGATGC	CCATCCGAAT	ATAACAGCAA	CAGCCCAAAC
951	TGCCCTTGCC	GTAGCAGAGG	CCGCAGGTAC	GGTTTGGCGC	GGTAAAAAAG
1001	TAGAACTTAA	CCCGACCAAA	TGGGATTGGG	TTAAAAATAC	CGGCTATAAA
1051	AAACCTGCTG	CCCGCCATAT	GCAGACTGTA	GATGGGGAGA	TGGCAGGGGG
1101	GAATAGACCG	CCTAAATCTA	TAACGTGCGA	AGGAAAAGCT	AATGCTGCAA
1151	CCTATCCTAA	GTTGGTTAAT	CAGCTAAATG	AGCAAAACTT	AAATAACATT
1201	GCGGCTCAAG	ATCCAAGATT	GAGTCTAGCT	ATTCATGAGG	GTAACAAAAA
1251	TTTTCCTAAT	GGAAGTGAAG	CTTATGAAGA	GGCAGATAGA	CTAGGTAAAA
1301	TTTGGGTTGG	TGAGGGTGCA	AGACAACTA	GTGGAGGCGG	ATGGTTAAGT
1351	AGAGATGGCA	CTCACAATA	TCGGCCACCA	ACAGAAAAAA	AATCACAATT
1401	TGCAACTACA	GGTATTCAAG	CAAATTTTGA	AACTTATACT	ATTGATTCAA
1451	ATGAAAAAAG	AAATAAAATT	AAAAATGGAC	ATTTAAATAT	TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

1	MNLPIQKFMM	LLAAAIISMLH	IPISHANGLD	ARLRDDMQAK	HYEPGGKYHL
51	FGNARGSVKN	RVCVQTFDA	TAVGPILPIT	HERTGFEGVI	GYETHFSGHG

101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
 151 GYPEPQGARD IYSYHIKGTSTKTKINTVPQ APFSDRWLKE NAGAASGFLS
 201 RADEAGKLIW ENDPDKNWRANRMDDIRGIV QGAVNPFLTGFQGVGIGAIT
 251 DSAVSPVTDTAQAQTLOGINDLGNLSPEAQLAAASLLQDSAFAVKDGINS
 301 ARQWADAHFNITATAQTALVAEAAAGTVWRGKKVELNPTKWDVVKNTGYKKPAARHMQTV
 351 KPAARHMQTV DGEMAGGNRP PKSTSEGKANAATYPKLVNQLNEQNLNNIAAQDPRLSLAIHEGKKNFP
 401 AAQDPRLSLAIHEGKKNFP GTATYEEADRLGKIWVGEGARQTSGGGWLSDGTRQYRPPTEKKSQFATTGFIQANFETY
 451 RDGTRQYRPPTEKKSQFATTGFIQANFETY IDSNEKRNKI KNHGLNIR*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

10	orf29ng-1.pep	10	20	30	40	50	60
		MNLP	IQKF	MMLL	AAAI	SMLH	IPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
	orf29-1	MNLP	IQKF	MMLL	FAAA	ISLL	QIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK
15	orf29ng-1.pep	70	80	90	100	110	120
		RVC	AVQT	FDATA	VGPI	LPI	THERTGFEGVIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG
	orf29-1	RVY	AVQT	FDATA	VGPI	LPI	THERTGFEGVIGYETHFSGHGHEVHSPFDHDSKSTSDFSG
20	orf29ng-1.pep	130	140	150	160	170	180
		GVD	GGFT	VYQL	LHRT	GSEI	HPADGYDGPQGGGYPEPQGARDIYSYHIKGTSTKTKINTVPQ
	orf29-1	GVD	GGFT	VYQL	LHRT	GSEI	HPEDGYDGPQGSDDYPPPGGARDIYSYVKGSTKTKTNIVPQ
25	orf29ng-1.pep	190	200	210	220	230	240
		APF	SDRW	LKEN	NAGA	ASGFL	SRAD
	orf29-1	APF	SDRW	LKEN	NAGA	ASGFF	SRAD
30	orf29ng-1.pep	250	260	270	280	290	300
		FQGV	GIGAIT	DSAV	SPVTD	TAAQ	QTLOGINDLGNLSPEAQLAAASLLQDSAFAVKDGINS
	orf29-1	FQGV	GIGAIT	DSAV	SPVTD	TAAQ	QTLOGINDLGKLSPEAQLAAASLLQDSAFAVKDGINS
35	orf29ng-1.pep	310	320	330	340	350	360
		ARQ	WADA	HPNIT	ATAQT	ALVA	EAAAGTVWRGKKVELNPTKWDVVKNTGYKKPAARHMQTV
	orf29-1	AKQ	WADA	HPNIT	ATAQT	ALSA	AAEAGTVWRGKKVELNPTKWDVVKNTGYKKPAARHMQTL
40	orf29ng-1.pep	370	380	390	400	410	419
		DGE	MAGG	NRPP	KSI-T	SEGK	ANAATYPKLVNQLNEQNLNNIAAQDPRLSLAIHEGKKNFP
	orf29-1	DGE	MAGG	NKPI	KSLP	NSAA	EKRKQNFEFNSNWSSASFDSVHKTLPNAPGILSPDKVKT
45	orf29ng-1.pep	420	430	440	450	460	470
		IGT	ATYEE	ADRL	GKIW	VGEG	ARQTSGGGWLSDGTRQYRPPTEKKSQFATTGFIQANFETY
50	orf29-1	RYT	SLDG	KITII	KDNEN	NYFRI	HDNSRKQYLDNSNGNAVKTGNLQKGQAKDYLQQQTHIRN
55		430	440	450	460	470	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

-148-

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 10 151 ATGAAGGAGA CAGAGGGGGC GTTCTCTCCA TTGGCTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
 15 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

20 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKETEGAFLEP LAILGGAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 101 PGGVGAAGKV VSEFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

	10	20	30	40
orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ			
30 orf30a	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP			
	10	20	30	40
	50	60		
orf30a	LXILGGAAGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI			
	70	80	90	100
	110	120		

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 40 151 ATGAAGGANA CAGNGGGGGC GTTCTCTCCA TTGGNTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
 45 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This encodes a protein having amino acid sequence <SEQ ID 178>:

50 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKXTXGAFLP LXILGGAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 101 PGXVGAAGKV VSEFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

orf30a.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

-149-

```

      ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
orf30-1  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60
5  orf30a.pep  LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI 120
      | |||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1  LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI 120
      |||||||||||||||||||||||||||||||||||||||||||||||||||||
10 orf30a.pep  KIGNNMRIAPFGNRTGHPGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR 180
      |||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1  KIGNNMRIAPFGNRTGHPGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR 180
      ||
orf30a.pep  FX
      ||
15 orf30-1  FX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

```

20 orf30.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ 42
      |||||||||||||||||||||||||||||:|||||||
orf30ng  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60

```

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

```

25      1  ATGAAAAAAC  AAATCACCGC  AGCCGTAATG  ATGCTGTCTA  TGATCGCCCC
51  CGCAATGGCA  AACGGATTGG  ACAATCAGGC  ATTTGAAGAC  CAAGTGTTC
101 ACACGCGGGC  AGATGCGCCG  ATGCAGTTGG  CGGAGCTTTC  TCAGAAGGAG
151 ATGAAGGAGA  CTGAAGGGGC  TTTTCTTCCA  TTGGCTATCT  TGGGTGGTGC
201 TGCCATTGGT  ATGTGGACAC  AGCATGGTTT  TAGTTATGCA  ACGACAGGCA
251 GACCAGCTTC  TGTTAGAGAT  GTTGCTGGCG  GATTAGGCGC  AATTCCTGGT
301 GATGTAGGTG  CTGCAGGAAA  GGTTGTTTCC  TTTGCTAAAT  ATGGACGTGA
30  351 GATTAAATC  GGCAATAATA  TGCGGATAGC  CCCTTCGGT  AATAGAACAG
401 GTCATCCTAT  TGGAAATTT  CCCCATTATC  ATCGTCGAGT  TACGGATAAT
451 ACGGGCAAGA  CTTTGCCTGG  ACAGGGAATT  GGTCTCATC  GCCCTGGGA
501 ATCAAAATCT  ACGGACAGAT  CATGGAAAAA  CCGCTTCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 180>:

```

35      1  MKKQITAAVM  MLSMIAPAMA  NGLDNQAFED  QVFHTRADAP  MQLAELSQKE
51  MKETEGAFLP  LAILGGAAIG  MWTQHGFSYA  TTGRPASVRD  VAGGLGAIPG
101 DVGAAGKVVS  FAKYGREIKI  GNNMRIAPFG  NRTGHPGKFP  PHYHRRVTDN
151 TGKTLPGQGI  GRHRPWESKS  TDRSWKNRF*

```

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

```

40      10      20      30      40      50      60
orf30ng.pep  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
      |||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1  MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
      10      20      30      40      50      60
45      70      80      90      100     110
orf30ng.pep  LAILGGAAIGMWTQHGFSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI
      |||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1  LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI
      70      80      90      100     110     120
50      120     130     140     150     160     170
orf30ng.pep  KIGNNMRIAPFGNRTGHPGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
      |||||||||||||||||||||||||||||||||||||||||||||||||||||
55 orf30-1  KIGNNMRIAPFGNRTGHPGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
      130     140     150     160     170     180
60 orf30ng.pep  180
      FX
      ||
orf30-1  FX

```

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.gonorrhoeae*

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N. gonorrhoeae*:

35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

	1	ATGAACAAAA	CCCTC1ATCG	TGTGATTTTC	AACCGCAAAC	GCGGTGCTGT
	51	GGTAGCTGTT	GCCGAAACCA	CCAAGCGCGA	AGGTAAGG	TGTGCCGATA
	101	GTGGTTCGGG	CAGCGTTTAT	GTGAAATCCG	TTTCTTTTAT	TCCTACTCAT
40	151	TCCAAAGCCT	TTTGTPTTTC	TGCATTAGCG	TTTTCTTTAT	GTTTGGCTTT
	201	GGGTACGGTC	AATATTGCTT	TTGCTGACGG	CATTATTACT	GATAAAGCTG
	251	CTCCTAAAA	CCAACAAGCC	ACGATTCTGC	AAACAGGTaa	cGGCATAACCG
	301	CAAGTCAATA	TTCAAAACCC	TACTTCGGCA	GGGGTTCTG	TTAATCAATA
	351	TGCCCAGTTT	GATGTGGGTA	ATCGCGGGGC	GATTTTAAAC	AACAGTCGCA
45	401	GCAACACCCA	AACACAGCTA	GGCGGTTGGA	TTCAAGGCCAA	TCCTTTGGTTG
	451	ACAAGGGGGG	AAGCAGGTGT	GGTTGTAAC	CAAAACAACA	GCAGCCATCC
	501	TTCACAACTG	AATGGGTATA	TTGAAGTGGG	TGGACAGCGT	GCAGAAGTCG
	551	TTATTGCCAA	TCCGGCAGGG	ATTGCAGTCA	ATGGTGGTGG	TTTTATCAAT
	601	GCTTCCCGTG	CCACTTTGAC	GACAGGCCAA	CCGCAATATC	AAGCAGGAGA
	651	CTTTAGCGGC	TTTAAGATAA	GGCAAGGCCAA	TGCTGTAAATC	GCCGGGACAG

1	MNKTLYRVIF	NRKRGAVVAV	AETTKREGKS	CADSGSGSVY	VKSVSFIPTH
51	SKAFPCFSALG	FSLCLALGTV	NIAFADGIIT	DKAAPKTQQA	TLQGTNGPI
101	QVNITQPTSA	GVSVNQYAQF	DVGNRGAILN	NSRNTQTOL	GGWIQGNPWL
151	TRGEARVVVN	QINSSHPSQL	NGYIEVGRR	AEVVIANPAG	IAVNGGGFIN
201	ASRATLTGQ	PQYQAGDFSG	FKIRQNAVI	AGHGLDARDT	DFTRILVCQQ
251	NHLDOYGRTS	RHS*			

orf31ng	96	GNGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIOGNPWLTRGE	154
HecA	45	GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L	104
Orf31ng	155	ARVVVNQINSSHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTGTGPQYQ	214
HecA	105	AAAILNEVVSFNNRSLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTGPQD	164
Orf31ng	215	-AGDFSGFKIROGNAVIAGHGLDARDTDF	242
HecA	165	AAGLSGLDVRGGDILIDGAGLDASRSDY	193

		10	20	30	40	50	60
25	orf31-1.pep	MNKTLYRVIFNRKRGAVVA	ETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS				
							:
	orf31ng	MNKTLYRVIFNRKRGAVVA	ETTKREGKSCADSGSGSVYKSVSFIPTH-----SKAFC				
		10	20	30	40	50	
30		70	80				
	orf31-1.pep	FSLLGFSLCLAVGTANIAFADGI					
				:			
	orf31ng	FSALGFSLCLALGTVNIAFADGIITDKAAPKTQQA	TILOTGNGIPIQVNIQTPTSAGVSVN				
		60	70	80	90	100	110

40

1	ATGAATACTC	CTCCTTTTGT	CTGTTGGATT	TTTTGCAAGG	TCATCGACAA
51	TTTCGGCGAC	ATCGGCGTTT	CGTGGCGGCT	CGCCCGTGTT	TTGCACCGCG
101	AACTCGGTTG	GCAGGTGCAT	TTGTGGACGG	ACGATGTGTC	CGCCTTGCGT
151	GCGCTTTGCC	CTGATTTGCC	CGATGTTCCC	TGCGTTTCAT	AGGATATTCA
201	TGTCCGCACT	TGGATTCCG	ATGCGGCAGA	TATTGATACC	GCG..

50 1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGC AAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGT TTGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCCT

-152-

151 GCGCTTTGCC CTGATTGGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA
 201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
 251 CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
 301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
 351 CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
 401 GTGTTCAAAA ATATTTTGGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
 451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
 501 CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
 551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
 601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
 651 CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
 701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
 751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
 801 CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCCTCT
 851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
 901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTGCGC
 951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
 1001 CACAACGCCT CGAATGTTGG CAAACCTGCG AACAACATCA AAACGGCTGG
 1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
 1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
 51 ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
 101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
 151 LIRERDYCEA VRFDTALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
 201 QAGSPMTLLL AGTQIIDLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
 251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
 301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLECW QTLQOHQNGW
 351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHOK IR*w

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N.*

meningitidis:

35 orf32.pep 10 20 30 40 50 60
 MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
 orf32a 10 20 30 40 50 60
 MNTPPFSAGXFECKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
 40 orf32.pep 70 80
 CVHQDIHVRTWHSDAADIDTA
 orf32a 70 80 90 100 110 120
 CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX

The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGT TGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
 151 GCGCTTTGCC CTGATTTGCC CGATGTTTNC TCGGTTTCATC AGGATATTCA
 201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
 251 NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
 301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
 351 CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
 401 GTGTTCAAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
 451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
 501 CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
 551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
 601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNNGGGGCGC ANATTATCGA
 651 CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
 701 CCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
 751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

10

15

25

30

35

40

45

22

55

60

orf32.pep MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTTDDVSALRALCPDLP 57

orf32ng MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP 60
 orf32.pep DVPCVHQDIHVRTWHSDAADIDTA 81
 5 orf32ng DVPFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDVVS
 51 ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVDAV IETFACDLP
 101 NVLNIIRRHK PLWLNWEYLS AEESENERLHL MSPQEGVQK YFWFMGFSEK
 151 SGGILIRERDY REAVRFDTEA LRRRLVLEPK NAPEWLLFGY RGDVWAKWLD
 201 MWQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
 251 PFVQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD
 301 KLHAFWDKAY GFYTPETASV HRLLSDDLNG GEALSATQRL ECGVL*

15 Further sequencing revealed the following DNA sequence <SEQ ID 195>:

1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTGTGCA AGGTCATCGA
 51 CAATTTCCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCACC
 101 GCGAACTCGG TTGGCAGGTG CATTGTGGA CGGACGACGT GTCCGCCTTG
 151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
 201 TCATGTCCCG ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
 251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
 301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAAATATT
 351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCGAG
 401 AGGCGGTTCA AAAATATTTT TGGTTATGTT GTTTCAGCGA AAAAAGCGGC
 25 GGTGTTGATC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
 501 AGCCCTGCGC CGCGCGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
 551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
 601 CAACAGGCGA GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
 651 CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAACGCC CTGCAAAATg
 701 aaggcgGTGT CTTTCagacg gcatccgTcC gccttGTCAA AAtcCCGTTT
 751 GTGCGCAAC AGGAcTTTCA CAAATGCTG CAcctgcCG ACTGCGCCGT
 801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCTT
 851 TTTTGTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAACTC
 901 CACGCTTTT GGGATAAGGC ATACGGCTTC TACAGCCCCG AAACCGCATC
 35 GGTGCGCCCG CTCCTTTTCG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
 1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
 1051 TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
 1101 CGCATCCGAA AAACCTCGCG CCTTTGTTTC AAAGCATCAA AAAATACGCT
 1151 AG

40 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDVVSAL
 51 RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVDAVIE TFACDLPENV
 101 LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
 45 151 GLIRERDYRE AVRFDTEALR RRLVLEPKNA PEWLLFGYRG DVWAKWLDMW
 201 QQAGSLMTLL LAGAIIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
 251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL
 301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLC WQTLQQHQNG
 351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

50 orf32-1.pep 10 20 30 40 50 59
 MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
 orf32ng-1 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
 55 10 20 30 40 50 60
 60 70 80 90 100 110 119
 orf32-1.pep PCVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
 orf32ng-1 PFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
 60 70 80 90 100 110 120
 120 130 140 150 160 170 179

-155-

5	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA
10	orf32-1.pep	SEWLLFGYRSDVWAKWLEMWROAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDFVQT
	orf32ng-1	PEWLLFGYRGDVWAKWLDWMQQAGSLMTLLLAGAQIIDSLKQSGVIPQNALQNEGGVFQT
15	orf32-1.pep	ASVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPPFWHIYPQDENVHLDKL
	orf32ng-1	ASVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRTQLAGKPPFWHIYPQDENVHLDKL
20	orf32-1.pep	HAFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATORLECWQTLQQHQNGWRQGAEDWSR
	orf32ng-1	HAFWDKAYGFYTPETASVHRRLSDDLNGGEALSATORLECWQTLQQHQNGWRQGAEDWSR
25	orf32-1.pep	YLFQGPSAPEKLAAFVSKHQKIRX
	orf32ng-1	YLFQGPSASEKLAAFVSKHQKIRX

- 30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 35 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

- 40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

1 ..TTGTTCTGC GTGTNAAAGT GGGGCGTTT TTCAGCAGTC CGGCGACGTG
51 GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
45 151 AGCCTGTGCG TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCAGC CTGTTGAGCA
251 ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301 CTCGGTTTCC CTGTCCCCGA TCGCGGTCG GTCATCGAAG GCCGTCTGAA
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401 TCGCCTGCTA NGGCATCCTG CCGCGCCTG..

- 50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

1 ..LFLRVKVGRE FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51 SLWLCTLLGM LVSLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

1	ATGTTGAATC	CATCCCGAAA	ACTGGTTGAG	CTGGTCCGTA	TTTTGGACGA
51	AGGCGGTTTT	ATTTTCAGCG	GCGATCCCGT	ACAGGCGACG	GAGGCTTTGC
101	GCCGCGTGGA	CGCGAGTACG	GAGGAAAAAA	TCATCCCGTCG	GCGCGAGATG
151	ATTGACAGGA	ACCGTATGCT	GCGGGAGACG	TTGGAACGTG	TGCGTGCGGG
201	GTCGTTCTGG	TTGTGGGTGG	TGGCGGCGAC	GTTTGCATTT	TTTACCGGTT
251	TTTCAGTCAC	TTATCTTCTA	ATGGACAATC	AGGGTCTGAA	TTTCTTTTTT
301	GTTTTCGGCG	GCGTGTGGG	CATGAATACG	CTGATGCTGG	CAGTATGGTT
351	GGCAATGTTG	TTCCCTGCGT	TGAAAGTGGG	GCGTTTTTTC	AGCAGTCCGG
401	CGACGTGGTT	TCGGGGCAAA	GACCTGTAA	ATCAGGCGGT	GTTGCGGCTG
451	TATGCGGACG	AGTGGCGGCA	ACCTTCGGTA	CGTTGAAAAA	TAGGCGCAAC
501	GTCGCACAGC	CTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTCGGTAT
551	TGTTGCTGCT	TTTGGTGGCG	CAATATACGT	TCAACTGGGA	AAGCAGCGTG
601	TTGAGCAATG	CCGCTTCGGT	ACGCGCGGTG	GAAATGTTGG	CATGGCTGCC
651	GTGAAACTC	GGTTTCCCTG	TCCCGGATGC	GCGGGCGGTC	ATCGAAAGCC
701	GTCGTAAACG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
751	GGCAGTATCG	CCATGCTACG	CATCCTGCGG	CGCCTGCTGG	TTTGGGTAGT
801	GTGTAAATC	CTTTTGAAAA	CAAGCGAAAA	CGGATTGGAT	TTGAAAAAGC
851	CCTATTATCA	GGCGGTCATC	CGCCGCTGGC	AGAACAAAAA	CACCGATTGG
901	GATACGCGTC	GGGAAACCGT	GTCCGCGGTT	TCACCGAAAA	TCATCTTGAA
951	CGATGCGCCG	AAATGGGCGG	TCATGCTGGA	GACCGAGTGG	CAGGACGCGG
1001	AATGTTTCGA	GGGCAGGCTG	GCGCAGGAAT	GGCTGGATAA	GGGCGTTGCC
1051	ACCAATCGGG	AACAGGTTGC	CGCGCTGGAG	ACAGAGCTGA	AGCAGAAACC
1101	GGCGCAACTG	CTTATCGGCG	TGCGCGCCCA	AACTGTGCCG	GACCGCGGGG
1151	TGTTGCGGCA	GATTGTCCGA	CTCTCGGAAG	CGGCGCAGGG	CGGCGCGGTG
1201	GTGCAGCTTT	TGGCGGAACA	GGGGCTTTCA	GACGACCTTT	CGGAAAAGCT
1251	GGACATTGGC	CTAACGCGC	TGGCCGAATG	CGGCGCGGGC	TGGCTTGAGC
1301	CTGACAGGGC	GGCGCAGGAA	GGGCGTTTGA	AAGACCAATA	A

1	MLNPSRKLVE	LVRILDEGGF	IFSGDPVQAT	EALRRVDGST	EKKIIRRAEM
51	IDRNRMLRET	LERVRAGSFW	LWVVAATFAF	FTGFSVTYLL	MDNQGLNFFL
101	VLACVLGMNT	LMLAVWLAMH	FLRVKVGRRF	SSPAWTFGRK	DPVNQAVLRL
151	YADEWRQPSV	RWKIGATSHS	LWLCTLLGML	VSVLLLLLLR	QYTFNWSTLL
201	LSNAASVRV	EWKLAVPLSKL	GFPVPDARAV	TEGRNLGNIA	DARAWSGLLV
251	GSIACYGILP	RLLAWVVCKI	LLKTSSEGLD	LEKPYQAVI	RRWQNKITDA
301	DTRRETVSVA	SPKIIINDAP	KWAVMLETEW	QDGEWEFGRL	AQEWALGGVA
351	TNREQAQVAE	TELKQKPAQL	LIGVRAQTVP	DRGVLRVQIV	LSEAAQGGAV
401	VOLLAEGLS	DDLSEKLEHW	RNALAECGAA	WLEPPDRAAO	GRLKDO*

Homology with a predicted ORF from *N.meningitidis* (strain A)

		10	20	30			
orf33.pep	LFLRVKVGRRFFSSPATWFRXKD	PVNQAVLR					
orf33a	LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGRFFSSPATWFRGKD	PVNQAVLR					
	90	100	110	120	130	140	
		40	50	60	70	80	90
orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVS	VLLLLVRQYTFNWESTLLSNAASVRA					
		:	:	:	:	:	
orf33a	LYADEWRXPSPSRWKIGATSHSLWLCTLLGMLVS	VLLLLVRQYTFNWESTLLGDSSSVRL					
	150	160	170	180	190	200	
		100	110	120	130	140	
orf33.pep	VEMLAWLP SKLGFPVPDARS	VIEGRNGNIADARAWSGLLVXSIA	CXGILPRL				
	: :						
orf33a	VEMLAWLP AKLGFPVPDARAVIEGRNGNIADARAWSGLLVGSIA	CYGILPRLLAWAVCK					
	210	220	230	240	250	260	
orf33a	ILXXTSENGLDLEKXXXXXXXIRRWQNKITDADRRETVSAVSPKIVLNDA	PKWAVMLETE					
	270	280	290	300	310	320	

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```

1  ATGTTGAATC CATCCCAGAA ACTGGTTGAG CTGGTCCGTA TTTTGAAGA
51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGGAAGATG
5  151 ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTTGCGTTT NTTACCGNTT
251 TTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
301 GTTTTGCGCG GCGTGNTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
10  351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
401 CGACGTGGTT TCGGGGCAAA GACCTGTCA ATCAGGCGGT GTTGGCGGTG
451 TATGCGGACG AGTGGCGGCG ACCTTCGTA CGTTGAAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551 TGTTCGTGCT TTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
15  601 TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
651 TGCGAAATCG GGTTCCTCCG TCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTGGTCCGGT GCTGCTGGTC
751 CGCAGTATCG CCTGCTACGG CATCTGCCG CGCTCTTGG CTGGGCGGCT
801 ATGCAAAATC CTNTGNAAA CAAGCGAAAA CGGCTTGGAT TTGAAAAGC
851 NNNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
20  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
951 CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GGCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCC GACCGCGGCG
25  1151 TGTTCGCGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTT CGGAAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301 CCGACAGAGC GCGCAGGAA GGCCGCTGTA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```

30  1  MLNPSRKIVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51  IDRNRMLRET LERVAGSFW LWVAAATFAF XTXFSVTYLL MDNQGLNFFL
101 VLAGVXGMNT LMLAVWLAML FLRVKVGREF SSPATWFRGK DPNVQAVLRL
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
35  201 LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 DTRRETSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401 VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRILKTNDRT*

```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```

40  10 20 30 40 50 60
orf33a.pep MLNPSRKIVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDNRMLRET
orf33-1 MLNPSRKIVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET
10 20 30 40 50 60
45  70 80 90 100 110 120
orf33a.pep LERVAGSFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
orf33-1 LERVAGSFWLWVAAATFAFTGFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
50  70 80 90 100 110 120
orf33a.pep FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
orf33-1 FLRVKVGREFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
130 140 150 160 170 180
55  130 140 150 160 170 180
orf33a.pep VSVLLLLLVRQYTFNWESTLLGSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRNLGNIA
orf33-1 VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
190 200 210 220 230 240
60  190 200 210 220 230 240
orf33a.pep DARAWSGLLVGSIACYGILPRLAWAVCKILXXTSENGLDLEKXXXXXXIIRRWQNKITDA
orf33-1 DARAWSGLLVGSIACYGILPRLAWAVCKILXXTSENGLDLEKXXXXXXIIRRWQNKITDA
250 260 270 280 290 300
65  250 260 270 280 290 300

```

-158-

5	orf33-1	DARAWSGLLVGS	250	260	270	280	290	300
	orf33a.pep	DTRRETSAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVAAANREQVAALE	310	320	330	340	350	360
10	orf33-1	TELKQKPAQLLIGVRAQTPDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW	370	380	390	400	410	420
	orf33a.pep	TELKQKPAQLLIGVRAQTPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW	370	380	390	400	410	420
15	orf33-1	RNALTECGAAWLEPDRAAQEGRLKNDRTX	430	440	450			
	orf33a.pep	RNALTECGAAWLEPDRAAQEGRLKNDRTX	430	440	450			
20	orf33-1	RNALAECGAAWLEPDRAAQEGRLKDQX	430	440				
	orf33a.pep	RNALAECGAAWLEPDRAAQEGRLKDQX	430	440				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae*:

25	orf33.pep	LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR	30
	orf33ng	LMDNQGLNFFLVLAGVLGMNTLMLAVWLATLFLRVKVGRRFFSSPATWFRGKGPVNQAVLR	100
30	orf33.pep	LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVVRQYTFNWESTLLSNAASVRA	90
	orf33ng	LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVVRQYTFNWESTLLSNAASVRA	160
35	orf33.pep	VEMLAWLPSKLGFPVPDARSVIEGRNLNGNIADARAWSGLLVXSIACXGILPRL	143
	orf33ng	VEMLAWLPSKLGFPVPDARAVIEGRNLNGNIADARAWSGLLVGSIVCYGILPRLAWVVCK	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

40	1	MIDRDRMLRD	TLERVAGSF	WLWVVASMM	FTAGFSGTYL	LMDNQGLNFF
	51	LVLAVGLGMN	TLMLAVWLAT	LFLRVKVGRR	FSSPATWFRG	KGPVNQAVLR
45	101	LYADQWRQPS	VRWKIGATAH	SLWLCTLLGM	LVSLLLLLV	RYTFNWEST
	151	LLSNAASVRA	VEMLAWLPSK	LGFPVPDARA	VIEGRNLNGNI	ADARAWSGLL
50	201	VGSIVCYGIL	PRLAWVVCK	ILLKTSSENG	DLEKTYQAV	IRRWQNKITD
	251	ADTRRETSA	VSPKIVLND	PKWALMLETE	WQDQWFEGR	LAQEWLDKGV
55	301	AANREQVAAL	ETELKQKPAQ	LLIGVRAQTV	PDRGVLRQIV	RLSEAAQGGG
	351	VVQLLAEQGL	SDDLSEKLEH	WRNALTECGA	AWLEPDRVAQ	EGRLKDQ*
60	1	ATGTTGaatC	CATCCCgaAA	ACTGgttgag	ctGgTCCgtA	Ttttgaataa
	51	aggggggtTTT	attttcagcg	gcgacccgtg	gcaggcgacg	gaggctttgc
65	101	gccgcgtgga	cggcAGTACG	GaggAaaaaa	tcttcgctcg	GGCGGAGAtg
	151	atcgACAGGg	accgtatggt	gcgggACaCg	TtggaacGTG	TGCGTGCGgg
70	201	gtcgtTctgG	TTATGGGTGG	TggtggCatC	gATGATGttt	aCCGCCGGAT
	251	TTTCAGgcac	ttatCttCTG	ATGGACaatC	AGGGGctGAA	TttCTTTTTTA
75	301	GTTTTggcgG	GAGTGTtggG	CATGaatacG	ctgATGCTGG	CAGTATGGtt
	351	gGCAACGTTG	TTCTGCGCG	TGAAAGTGGG	ACGGTTTTTC	AGCAGTCCGG
80	401	CGACGTGGT	TCGGGGCAAA	GGCCCTGTAA	ATCAGGCGGT	GTTGCGGCTG
	451	TATGCGGACC	AGTGGCGGCA	ACCTTCGGTA	CGATGGAAAA	TAGGCGCAAC
85	501	GGCGCACAGC	TTGTGGCTCT	GCACGCTGCT	CGGAATGCTG	GTGTGCGGTAT
	551	TGCTGTGCT	TTTGGTGGG	CAATATACGT	TCAACTGGGA	AAGCACGCTG
90	601	TTGAGCAATG	CCGCTTCGGT	ACGCGCGGTG	GAAATGTTGG	CATGGCTGCC
	651	GTCGAAACTC	GGTTTCCCTG	TCCCCGATGC	GCGGGCGGTC	ATCGAAGGTC
95	701	GTCTGAACCG	CAATATTGCC	GATGCGCGGG	CTTGGTCGGG	GCTGCTGGTC
	751	GGCAGTATCG	TCTGCTACGG	CATCCTGCCG	CGCCTCTTGG	CTTGGGTAGT

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

-159-

801 GTGTAATC CTTTGA AAA CAAGCGAAAA CGGattgGAT TTGGA AAAA
 851 CCTATTATCA GCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901 GATACGCGTC GGGAAACCGT GTCCCGCGTT TCGCgaAAA TCGTCTTGAA
 5 951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
 1051 GCCAATCGGG AACAGTTGCG CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
 1101 GGCCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 10 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGT GGGCAGGAA GGCCGTTTGA AAGACCAATA A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRAEM
 51 IDRDRMLRDT LERVAGSFW LWVVASMMF TAGFSGYLL MDNQGLNFFL
 15 101 VLAGVLMNT LMLAVLATL FLRVKVRFF SSPATWFRGK GPVNQAVLRL
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
 251 GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYQAVI RRWQNKITDA
 301 DTRRETVS AV SPKIVLNDAP KVALMLET EW QDGQWFEGR LAQEWLDKGVA
 20 351 ANREQVAAL TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKQD*

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

		10	20	30	40	50	60
25	orf33-1.pep	MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRMLRET					
	orf33ng-1	MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEEKIFRAEMIDRDRMLRDT					
		10	20	30	40	50	60
30	orf33-1.pep	LERVRAGSFWLWVVAATFAFFTFGSVTYLLMDNQGLNFFLVLAVGLMNTLMLAVWLAML					
	orf33ng-1	LERVRAGSFWLWVVASMMFTAGFSGYLLMDNQGLNFFLVLAVGLMNTLMLAVWLATL					
		70	80	90	100	110	120
35	orf33-1.pep	FLRVKVRFFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML					
	orf33ng-1	FLRVKVRFFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSLWLCTLLGML					
		130	140	150	160	170	180
40	orf33-1.pep	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
	orf33ng-1	VSVLLLLLVRQYTFNWESTLLSNAASVRVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA					
		190	200	210	220	230	240
45	orf33-1.pep	DARAWSGLLVGSIACYGILPRLLAWVVCKILLKTSENGLDLEKPYQAVIRRWQNKITDA					
	orf33ng-1	DARAWSGLLVGSIACYGILPRLLAWVVCKILLKTSENGLDLEKTYQAVIRRWQNKITDA					
		250	260	270	280	290	300
50	orf33-1.pep	DTRRETVS AVSPKIVLNDAPKWVMLETEWQDGQWFEGR LAQEWLDKGVA TNREQVAAL					
	orf33ng-1	DTRRETVS AVSPKIVLNDAPKWVMLETEWQDGQWFEGR LAQEWLDKGVA ANREQVAAL					
		310	320	330	340	350	360
55	orf33-1.pep	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
	orf33ng-1	TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW					
		370	380	390	400	410	420
60	orf33-1.pep	RNALAECGA AWLEPDRAAQEGRLKQDX					
	orf33ng-1						
		430	440				
65	orf33-1.pep						

orf33ng-1 RNALTECGAAWLEPDRVAQEGRLKDQX
430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
      51  CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT.GAGTGC
     101  CCTGTTTTTC GGGTGTCTTCT TTTCGGGGTT CGGGACGGGG GACGTTTGTC
     151  GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTC TTCC.GGCGT
     201  CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTG CG..GTTTGA
     251  CCCGGTTTTT CTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGTTTTCG
     15  301  TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
     351  GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTTC
     401  GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
     451  GTCC..

```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVG GLVWFSLVGS XECACFSGVS FRGSGRGTFFV
     51  GSTGVSLSVF SACVXGVVRL PVGLSCVGR LXXLTRFFLGA AGDVILLPLS
     101  SVPSGCAGSD EAAWWSGWA ASCPTTFFGS QNSVSRGLSV CCGSA*RVLS
     151  S..

```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
     51  GCCGGGTCAG AATAGGTTGT CCAGAATTTC TTTATGGGGT TTGGGCGGGC
     101  TGTTTTTCGG GGTGTCCGGT TTGTATAGGT TTTCTTTGGG CGTTTCTTTG
     151  GGTGCGCCCT GTTTTTCGGG TGTTCCTTTT CGGGGTTCGG GACGGGGGAC
     201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTTGTGTTC
     30  251  CGGCGTCGTC CGGCTGCCTG TCGGTTTGGT CTGTGTCCGG AGGTGCGGT
     301  TTGACCCGGT TTTTCTTGGG TCGCGCAGGG GACGGCAGTC CGCTGCCGCT
     351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
     401  GTTCGGGTTG GCGCGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAA
     451  TCGGTTTCGC GGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
     35  501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
     551  TTGCGCGCAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
     601  AGCCTGAAGG GTTTGTTCGG TTTTCTTGCC ATTTTGATTG TGCTTTTGGG
     651  GTGTCGGGCA ATGCCGTCTG AAGGCGGTTC AGACGGCATT GCCGAGTCAG
     701  CGTTGGACGT AGTTTGTGTA GAGGGTGATG ACTTTTGTGA CGCCGACGGT
     40  751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
     801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
     851  GTGCGGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTCCA
     901  AGTGTCCGGC GCGATGTGCG CCGCAGTGCG CGGCAGGGAG GCGACGGTAA
     951  TATAGTTGTA CACGCCTTCG GCGGCTGTT CGGAACGTGC AATCTGACCG
     45  1001  ACGAAGTGT TTTCCGCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
     1051  GCGGTTGTAG CCGACGACGG AGATTGGGG CGTGTAGCCT TTGGTTTGGT
     1101  TGTTTTGGCG CAGATAGGAA CCGGCGGTGG TTTGATACG CAACGCCATA
     1151  ACGTTGTGCG CGGTGTGCGC GCCGTTGGTT CGGCGGTGCA CGGCGGATTT
     1201  CGCGCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
     50  1251  CAAGGCTGAA AATGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
     1301  TTCATCGGGT GCTTCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
     1351  CATGCCGTCT GA

```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MPMFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLVGSL
     51  GCACFSGVSF RSGRGTFFV STGVSLSVFS ACVPASSGCL SV*AVSAGCG
     101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTFFGSQN
     151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV

```

5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 *meningitidis:*

[illegible]

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

35	1	ATGATGATNC	CGTTNATAAT	GCTTCCTTGG	ATTGCGGGTG	TGCCTGCCCT
	51	GCCGGGTGAG	AAGAGGTTGT	CGAGAANTTC	TTTATGGCGT	TTAGGCGGCN
	101	TGTTTTTCGG	GGTGTCCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTNTT
	151	TCTTTGGGTG	TTTCTNTGGG	CTGTGCCTGT	TTTTCGGGTG	TTTCTTTTCG
40	201	GGGTTTCGGG	CCGGGGACGT	TTGTGGGCAG	TACNCGGGTT	TCTTTGAGTG
	251	TGTTTTGAGC	TTGTGCTCCG	CGGTGCTCCG	GCTGCCTGTC	GGTTTTNAGT
	301	GTGTGCGCAG	GTTGCGGTTT	GACCCGGNTT	TTCTTTNGGTG	CGGCAGGGGA
	351	CGGCAGTCCG	CTCGCGCTTT	CGTCTGTGCC	GTCGGGCTGT	GCGGGTGCGG
45	401	ATGAGGAGGC	TGNTNGTGT	TCGGGTTGGG	GGGCATCTTG	TCCGACTACG
	451	CCGTTTGCCA	GCCAGAATTC	GGTTTCGCGG	GGGCTGTGCG	TGTGTTGCGG
	501	TTGCGTNTGG	AGGGTTTGTG	CNCCGTTCCG	TGNGAATGTG	CTGACGATGC
	551	CTATTGCCAA	TGCGCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCCGCT
50	601	ATCAGGAGTT	TGGGGGTCAG	CCTGAAGGGT	TTGTTTCNGT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCGGGCAAT	GCCGTCTGAA	GGCGGTTTCAG
	701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTNGGTAGA	GGGTGATGAC
	751	TTTTTGTAGC	CCGACGTGGG	TGCTGACTTT	TTGGGTAATC	TGCGCCTGTT
55	801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACGTTGCC	GTAGGTAACG
	851	ATTTTGAGCG	GCGCCTGTGT	GCGGGGGCTG	ATGCCCAACA	GCGTGGGCGG
	901	GACTTTGGAT	GTGTTCCTAA	TGTCGCGCGG	GATGTCGCGG	GCAGTGCGGG
	951	GACGGGAGGC	GACGGTAATG	TANTTGTACA	GCGCTTCGCG	GGCCTGTTCG
60	1001	GAACGTGCAA	TCTGACCGAC	GAACGTGTTT	TGCGCTTCGG	TGGCGACTTG
	1051	TCCGAGCAGC	AGCAGGTGGC	GGTTGTAGCC	GACAACGGAG	ATTTGGGGCG
	1101	TGTANCTTTA	GGTTTGGTTG	TTTTGGCGCA	GATAGGAGCG	GGCGGTGGTT
	1151	TGATACACGA	GCGCCATTAC	GTGTGCTGCG	GTTNGGCGCG	GTTGGGTTTCG
	1201	GCGGTCGACG	GCGGATTTTC	CGCCGACCGC	CGCGCCGCCG	ACGACTGCGC
	1251	TGACGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGCGCGCAGT	CAGGGTGCGG
	1301	ACGGTGTGCG	GTTTGGGTTT	CATCGGGTGC	TTCTTTTCTT	GGGCGTTTCA
	1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence <SEQ ID 212>:

5

10

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

15

20

25

30

35

40

45

50

55

60

Homology with a predicted ORF from *N.gonorrhoeae*

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae*:

OKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC 35

15 The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

This encodes a protein having amino acid sequence <SEQ ID 214>:

55 ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

```

      10       20       30       40       50
orf34-1.pep M M P F I M L P W I A G V P A V P G Q N R L S R I S L W G L G G V F F G V S G L V W F S L G V S ----- L G C A C
||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
60 orf34ng   M M P F I M L P W I A G V P A V P G Q K R L S R I S L W G L A G V F F G V S G L V W F S L G V S F S L G V S L G C A C
      10       20       30       40       50       60

      60       70       80       90      100      110
orf34-1.pep F S G V S F R G S G R G T F V G S T G V S L S V F S A C V P A S S G C L S V X A V S A G C G L T R F F L G A A G D G S P
||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
65 orf34ng   F S G V S F R G S G W A F V G S T G V S L S V F S A C V P V P V N E S A A R A A S E G R G L T R F F L G A A G D G S P

```

-164-

		70	80	90	100	110	120
		120	130	140	150	160	170
5	orf34-1.pep	LPLSSVPSGCAGSDEAAWCSGWAASCP	TTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV				
	orf34ng	LPLSSVPSGCAGSDEAAWCSGWAASCP	TAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV				
		130	140	150	160	170	180
10	orf34-1.pep	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
	orf34ng	LTMPANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		180	190	200	210	220	230
		190	200	210	220	230	240
15	orf34-1.pep	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
	orf34ng	LDVVLVEGNDFLYADGGADFLGNLRLFFGGEDAHNVGYIAVGNDFDARLCGGADAQQRGA					
		240	250	260	270	280	290
		250	260	270	280	290	300
20	orf34-1.pep	DFGCVPSVAGDVAGSARQGGDGNVVFHAFGGFLFGTCNLTDDELFFAFGGDLSEQQVAVVA					
	orf34ng	DFGRVPSVAGDVARSARQGGDGNVVVYAFGGFLFGTCNLTDDELFFAFGGDLSEQQVAVVA					
		300	310	320	330	340	350
25		310	320	330	340	350	360
	orf34-1.pep	DDGDLGRVAFGLVVLAAQIGTGGGFDTQRHNVVVLGRAGGSVDGGFRADGGASDYCADAA					
	orf34ng	DDGDLGRVAFGLVVLAAQVGTGGGFDTQRHNVVIGLRAGGSVDGFCADGGPADDCAEAA					
		360	370	380	390	400	410
30		370	380	390	400	410	420
	orf34-1.pep	AKGKAENGNGQAGDGVRFGRVLPFLGVSDGIALRHAVX					
	orf34ng	AEGKAEDGGNGQAGDGVWFGFHRGLPFLGVSDGIALRHAVX					
		420	430	440	450		
35		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

45      1  ATGAAACCT TCTTCAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
     151  CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG
     201  AGAAAAAAGG CTACACCGTC AACTGGTTCG AGTTTACCGA CTATGTACGC
     251  CCGAATCTGG CATTGGCTGA GGGCGAGTTG

```

50 This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

      1  MKTFFKTLA AALALILAAC G.QKDSAPAA SASAADNGA AKKEIVFGTT
     51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

55      1  ATGAAACCT TCTTCAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
      51  CGCCGCCTGC GGCGGTCAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
     101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
     151  GTCGGCGACT TCGCGGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
     201  GAAAAAAGGC TACACCGTCA AACTGGTTCG GTTTACCGAC TATGTACGCC

```

5
10
20
25
30
35
40
45
50
55
60
65
70
75
80
85

251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG
351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
451 CCGTCCAAC TCGCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGGAT
501 CAACTCAA GACGGCATCA ATCCGTTGAC CGCATCCAA GCGGACATCG
551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CCGCAACTG
601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

15
20
25
30
35
40
45
50
55
60
65
70
75
80
85

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRENLALE GELDINVFQH
101 KPYLDDFKKE HNLDTIEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND
151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKN KIVELEAAQL
201 PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNW AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEQYKSPAA WNEGAAK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180

orf4.pep
orf4a
orf4.pep
orf4a
orf4a

MKTFFKTLA AALALILAACG-QKDSAPAASASAAADNGA AKKEIVFGTTVGDFGDMVKE
MKTFFKTLA AALALILAACGGQKDSAPAASASAAADNGA AXKEIVFGTTVGDFGDMVKE
QIQAELEKKG YTVKLVEFTD YVRENLALE GELDINVFQH
XIQPELEKKG YTVKLVEFTD YVRENLALE GELDINVFQH
VPTAPLGLYP GKLSLEEVK DGSTVSAPND PXXFXRVLVMDLDELGXIKLKDIXXXXXXX

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180

1 ATGAAAACCT TCTTCAAAC CCTTCCGCC GCGCACTCG CGCTCATCCT
51 CGCGCGCTGC GCGGTCAA AAGATAGCGC GCCGCGCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAANAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGCGGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA
201 GAAAAAGGC TACACCGTCA AACTGGTCGA GTNTACCGAC TATGTGCGCN
251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTNCAACAC
301 ANACNCTATC TTGACGACTN CAAAAANAA CACAATCTGG ACATCACCNN
351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401 AATCGCTGGA NNAAGTCAA GANGCAGCA CCGTATCCGC GCCCAACGAC
451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTNGAT
501 CAACTCAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANN
551 NNNNANNNT NNNNNNNNN NNNNNCNCG NNNNNNNAN NNNNNNNNN
601 NCGNNTNNN NNGCNNTNN NNNNTNNN NNCNCCNNN NNNNTNNNN
651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AXKEIVFGTT

```

51  VGDFGDMVKE XIQPELEKKG YTVKLVEXTD YVRXNLALAE GELDINXQH
101 XXYLDDXKKX HNLDTXVKQ VPTAPLGLYP GKLSLXXVK XGSTVSAPND
151 PXXFXRVLMV LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

```

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

```

1  ATGAAACCT TCTTCAAAAC CCTTCCGCC GCCGCACTCG CGCTCATCCT
51  CGCCGCTGCG GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
101 CGCCGCCGCG CAACGGCGCG GCGAAAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
201 GAAAAAGGC TACACCGTCA AACTGGTCGA GTTACCGAC TATGTGCGCC
251 CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCCAACAC
301 AAACCCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCAG
351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCGG GGCAAGCTGA
401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCAACGAC
451 CGGTCCAAC TCGCCCGCGT CTGGTGATG CTCGACGAAC TGGGTTGGAT
501 CAACTCAAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
551 CCGAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601 CGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
851 GCGCAGCCAA ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

```

1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT
51  VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINXQH
101 KPYLDDFKKE HNLDTVEFQ VPTAPLGLYP GKLSLEEVEK DGSTVSAPND
151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLNKI KIVELEAAQL
201 PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

```

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

```

35  orf4a-1      10      20      30      40      50      60
      MKTFFKTLSA AALALILAACGGQKDSAPAA SAAADNGAAKKEIVFGTTVGDFGDMVKE
      orf4-1      10      20      30      40      50      60
      MKTFFKTLSA AALALILAACGGQKDSAPAA SAAADNGAAKKEIVFGTTVGDFGDMVKE

40  orf4a-1      70      80      90      100     110     120
      QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINXQH KPYLDDFKKE HNLDTVEFQ
      orf4-1      70      80      90      100     110     120
      QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINXQH KPYLDDFKKE HNLDTVEFQ

45  orf4a-1      130     140     150     160     170     180
      VPTAPLGLYP GKLSLEEVEK DGSTVSAPND PSNFARVLV MLDELGWIKL KDGINPLTASK
      orf4-1      130     140     150     160     170     180
      VPTAPLGLYP GKLSLEEVEK DGSTVSAPND PSNFARVLV MLDELGWIKL KDGINPLTASK

50  orf4a-1      190     200     210     220     230     240
      ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS
      orf4-1      190     200     210     220     230     240
      ADIAENLNKI KIVELEAAQL PRSRADVDF VVNGNYAISS GMKLTEALFQ EPSFAYVNWS

55  orf4a-1      250     260     270     280
      AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX
      orf4-1      250     260     270     280
      AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX

```

		10	20	
5	lip2.pasha	MNFKLLGVALVSALALTACKDEKAQAP----		
		: : : :		
	ORF4	VXTPNPDGRTPCPSFLFETATTSGENMKTFKTL\$A\$AAL--ALILAACGFKKATRP\$PHL		
		110	120	130 140 150
10	lip2.pasha	-ATTAKTENKAPLKVGVMGTGPEAQMTEVAVKIAKEKYGLDVELVQFT EYTQPNAALH\$SKD		
		: :: : : :: :: :: : : : : : :		
	ORF4	LPPPTTARRKKEIVFGTTVGDFGD\$MVKEIQIQAELEKKGYTVKLVEFTDYVRPNLALEGE		
		160	170	180 190 200 210
15	lip2.pasha	LDANAFQTVPYLEQEVDKDRGYKLAIIGNTLVWPIAAYSKIKNIS ELKD GATVAIPNNAS		
	ORF4	\$.....		

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

		10	20	30
25	orf4nm.pep	MKTFFKTL	SAAALALI	AACGXQKDSAPAA
			:	
	orf4ng	RANAVXTPNP	DGRTPCLSFLFETATTSGENMKTFFKTLSTASLALI	AACGGQKDSAPAA
		200	210	220
		230	240	250
		40	50	60
30	orf4nm.pep	SASA-AADNGA	AKKEIVFGTTVGDFGDMVKEIQAELEKKGYTVKLVEFTDYVRPNLALA	
		: :		
	orf4ng	SAAAPSADN	GAAKKEIVFGTTVGDFGDMVKEIQAELEKKGYTVKLVEFTDYVRPNLALA	
		260	270	280
		290	300	310
35		90		
	orf4nm.pep	EGEL	.	
	orf4ng	EGELDINVQH	KPYLD DFK EHN LDITEAF Q V T P A L G L Y P G K L S L E E V K D G S T V S A P N	
		320	330	340
		350	360	370

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

	1	MKTFFKTLST	ASLALILAAC	GGQKDSAPAA	SAAAPSADNG	AAKKEIVFGT
	51	TVGDFGDMVK	EHQIAELEKK	GYTVKVLVEFT	DYVRLPNLAA	EGELDINVFT
45	101	HKPYLDDFKK	EQNPATLDEAF	QVPTAPLGLY	PGKLSLEEV	KDGSTVSAFN
	151	DPSNFARALV	MLNELGWIKL	KDGINPLTAS	KADIAENLKN	IKIVELEAAQ
	201	LPRSRADVDF	AVVNGNYAIS	SGMKLLEALF	QEPSFYVNW	SAVKTADKDS
	251	OWLKDVTYAY	NSDAFKAYAH	KRFEGYKYPA	AWNEGAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

50	1	atgAAAACCT	TCTTCAAAAC	ccttttcgcgc	gccgcaCTCG	CGCTCATCCT
	51	CGCAGCCTGc	ggCggtcaAA	AAGACAGCGC	GCCCgcagcc	tctgcCGCCG
	101	CCCCTTCTGC	CGATAACGgc	gCgGCGAAAA	AAGAAAtcgt	ctTCGGCAGC
	151	Accgtggggc	acttcggcgA	TatggTCAAA	GAACAAATCC	AagcCGAgct
55	201	gGAGAAAAAA	GgctACACcg	tcAAattggt	cgaatttacc	gactatgtGC
	251	gCCCGAATCT	GGCATTGGCG	GAGGGCGAGT	TGGACATCAA	CGTCTTCCAA
	301	CACAACCCCT	ATCTTGACGA	TTTCAAAAAA	GAACACAACC	TGGACATCAC
	351	CGAAGCCTTC	CAAAGTGCCG	CCGCGCCTTT	GGGACTGTAT	CCGGGCCAAAC
60	401	TGAAATCGCT	GGAAGAAGTC	AAAGACGGCA	GCACCGTATC	CGCGCCCAac
	451	gACcgcTACC	ACTTTCGACG	CGCCTTGGTG	ATGCTGAACG	AACTGGGTTC
	501	GATCAAACTC	AAAGACGGCA	TCAATCCCGT	GACCGCATCC	AAAGCCGACA
	551	TCGCGGAAAA	CTGAAAAAAC	ATCAAAATCG	TCGAGCTTGA	AGCCGCGCAA

-168-

601 CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
 651 CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
 701 GCTTTGCCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
 751 CAATGGGCTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
 801 CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
 851 AAGGCGCAGC CAAATAA

This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SAAAPSAADNG AAKKEIVFGT
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVQ
 101 HKPYLDDEFK EHNLDITEAF QVPTAPLGLY PGKLKSLLEV KDGSTVSAFN
 151 DPSNFARALV MLNELGWIKL KGINPLTAS KADIAENLKN IKIVELEAAQ
 201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
 251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYP AAWNEGAAK*

This shows 97.6% identity in 288 aa overlap with ORF4-1:

15	orf4-1.pep	10	20	30	40	50	59
		MKTFFKTLA AALALILAACGGQKDSAPAAASASA-AADNGAAKKEIVFGTTVGDFGDMVK					
	orf4ng-1	MKTFFKTLA AALALILAACGGQKDSAPAAASAAAPSAADNGAAKKEIVFGTTVGDFGDMVK					
20		10	20	30	40	50	60
	orf4-1.pep	60	70	80	90	100	119
		EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDEFKKEHNLDITEVF					
	orf4ng-1	EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDEFKKEHNLDITEAF					
25		70	80	90	100	110	120
	orf4-1.pep	120	130	140	150	160	179
		QVPTAPLGLYPGKLKSLLEVVDGSTVSAPNDPSNFARVLVMDLGLWIKLKDGINPLTAS					
	orf4ng-1	QVPTAPLGLYPGKLKSLLEVVDGSTVSAPNDPSNFARALVMDLGLWIKLKDGINPLTAS					
30		130	140	150	160	170	180
	orf4-1.pep	180	190	200	210	220	239
		KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW					
	orf4ng-1	KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW					
35		190	200	210	220	230	240
	orf4-1.pep	240	250	260	270	280	
		SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPA AAWNEGAAKX					
	orf4ng-1	SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYP AAWNEGAAKX					
40		250	260	270	280		

In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

ID LIP2_PASHA STANDARD; PRT; 276 AA.
 AC Q08869;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE 28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .
 SCORES Init1: 279 Initn: 416 Opt: 494
 Smith-Waterman score: 494; 36.0% identity in 275 aa overlap

50	orf4ng-1.pep	10	20	30	40	50
		MKTFFKTLA AAL--ALILAACGGQKDSAPAAASAAAPSAADNGAAKKEIVFGTTVGDFGDM				
	lip2_pasha	MNFKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM				
60		10	20	30	40	50
	orf4ng-1.pep	60	70	80	90	110
		VKEIQAELEKKGYTVKLVEFTDYVRPNLALAEGLDINVQHKPYLDDEFKKEHNLDITE				
		:: :: : : : : : : : : : : ::				

-169-

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTQPNAAALHSKDLNANAFQTPVYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	120 130 140 150 160 170 AFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT						
	lip2_pasha	IGNTLVWPAAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDPKN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	180 190 200 210 220 230 ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA						
	lip2_pasha	ATENDIENPKNIKIVQADTSLLRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKSDSP	180	190	200	210	220	230
15	orf4ng-1.pep	240 250 260 270 280 289 YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEQYKYPAAWNEGAAKX						
20	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFTQTEEVFQELKLFNGGVVKGW	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

	1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
	51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
40	101	GTTTTATGGC	TGAACCAGCA	TTATTTCAC	GGCAACCTCC	TCTTCTACCT
	151	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG	CTGGGCGGCG	GTCGGCAAAA
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCGC	CGCCAAACTG	CTGCCGCTGA
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
45	401	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAACG
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
	501	AAAGCCGCAG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC
	551	GCCATGATGG	AAGCCATGCA	GCACGCCAC	CGTAAATCG	TCAACACCAC
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC	AAACTCAACG

-170-

5 651 GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....
 701 GC AGACACGCCC GCCGCATCCG
 751 CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC
 801 ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA
 851 ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA
 901 TGCCACGAA CGCCAACACC TGCGCCAAG CCGCTTGA

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10 1PRRP RHAPVSRGDL LQGGGYARH GHRAGRGFGR FMAEPALFPR
 51 QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQT
 101 HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
 151 AHDPTPRGE HGENAPNQRT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP
 201 XNRQHHRAAP DHRQAAISQ TQRQNPAAAX PPLHTAPN... Q
 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHFAATH
 301 PPQMAGCPRT PTPAPKPA*

15 Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N.gonorrhoeae*:

25 orf8ng 1 MDRDDLRLRRPHAPVPRRDLLQGGGYARYGHRAGRGFGRFMAEPALFPR 50
 orf8.pep 1PRRP RHAPVSRGDL LQGGGYARHGHRAGRGFGRFMAEPALFPR 44
 orf8ng 51 QPLLPDHRHGKRTGRLGGGRQKRLRPYVGADDVHAHRRQRQRMARQRP 100
 orf8.pep 45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94
 30 orf8ng 101 DARDERPHRRRHRCRRQTAAAEIHTDVAFHACRQPGRLQONDCRNQQRQ 150
 orf8.pep 95 HARHERPHRRGHRHRRRQTAAEIHTDVAFHACRQPGRMQONDCRNQQRQ 144
 35 orf8ng 151 AYDARTFGAEYQONAPNQRTHGQKPQPPRRHIGRKPQPLHDGSHAARPP 200
 orf8.pep 145 AHDPTPRGEHGENAPNQRTHGQKPQPSRRHIGRKLHQPRLHDGSHAARPP 194
 40 orf8ng 201 QNRQHHRAAPDHRQAAISQTQRQNPAAAPPLHTAPNRPATNRRPHQRQ 250
 orf8.pep 195 XNRQHHRAAPDHRQAAISQTQRQNPAAAXPPLHTAPN.....Q 244
 orf8ng 251 TRPPHPRHRHQPRTGSPRRTPPLPMAGFPLAQHRYASGNFRPRHPPATH 300
 45 orf8.pep 245 TRPPHPRHRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHFAATH 294
 orf8ng 301 PPQMAGCPRTPTPAPKPA* 319
 orf8.pep 295 PPQMAGCPRTPTPAPKPA* 313

50 The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

55 1 MDRDDLRLRRP RHAPVPRRD LQGGGYARY GHRAGRGFGR FMAEPALFPR
 51 QPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
 101 DARDERPHRR RHRHRCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
 151 AYDARTFGAE YQONAPNQRT HGQKPQPPRR HIGRKPQPL HDGSHAARPP

201 QNRQHHRAAP DHRROAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ
 251 TRPPHPRHRH HQPRTGSPRR TPPLPMAGFP LAQHGYASGN FRPRHPPATH
 301 PPQMAGCPRT PTPAPKPA*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

10      1  ..GAAATCAGCC TGC GGTCCGA CNACAGGCCG GTTTCCTGN CGAAGCGGCCG
      51  GGATTCGGAA CGT'TTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
      101  GGGCGTGGGT GGAACACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
      151  CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
      201  TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
      251  TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
      301  GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAGAAGC ACGGT'TCCGA
      351  CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTAGCCGC AACGCCTGCG
      401  TCGTCGT CAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
      451  GGACATTATC TCGGAGA.GG AACCATCATG CCCGTTTCC ACCTGATGAA
      501  AGAATCGCTC GCCGTCCGAA CCGCAACCT CAACCGGCAC GCCGGTAAGC
      551  GTTATCCTTT CCCGACCGG..

```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

      1  ..EISLRSDXRP VSVXKRDRSE RFLLLDGGNS RLKWAUVENG TFATVGSAPY
      51  RDLSPGLAEW AEKADGNVRI VGCAVCGEFK KAQVQQLAR KIEWLPSSAQ
      101  AXGIRNHRYH PEEHGS DRWF NALGSRRFSR NACVVVSCGT AVTVDALTDD
      151  GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..

```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

      1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
      51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
      101  CGCAGCAGCT CAACGGT'TT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
      151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGCGCGT
      201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTGC GGT'TTTCAGA
      251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
      301  GCGCGGATTG CGCCGACAAA GCGCACAAA ACCATATGCG TGACCCACCT
      351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
      401  GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
      451  GAGTTGGGTT CGCTGTCGCC TGTGCGGCA GTGGCGTGC GCGCGCGCTT
      501  TCGCGTTTGA GGT'TTGGATG TGCAGATTAA GTGGCCCAAT GATT'TGGTTG
      551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
      601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
      651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
      701  GCGCGGGCAA TGCCGATGCC GCCGTGCTGC TGGAACGCT GTTGGTGGAA
      751  CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGGATT'TG CGCCTTTTGT
      801  GCGGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
      851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
      901  CAAGCGGTTT TGCATTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
      951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
      1001  GCGGGGATTC GGAACGTTT CTGCTGTTGG ACGCGGCGCA CAGCCGGCTC
      1051  AAGTGGGCGT GGTGGAATAA CCGCACGTTC GCAACCGTCG GTAGCGCGCC
      1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGCGGATG
      1151  GAAATGTCCG CATCGTCGGT TGCCTGTGTG GCGGAGAATT CAAAAAGGCA
      1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
      1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
      1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
      1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCAG
      1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
      1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
      1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
      1551  GGATGCGGTT TCGCGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
      1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG

```

5	1	MTVLKLSHWR	VLAELADGLP	QHVSQRLARMA	DMKPPQQLNGF	WQOMPAAHIG
	51	LLRQHDGYWR	LVRPLAVFDA	EGLRELGERS	GFQTALKHEC	ASSNDEILEL
	101	ARIAPDKAHK	TICVTHLQSK	GRGRQGRKWS	HLRGELCMFS	FGWVFDRTQY
	151	ELGSLSPVA	ACVRRALSRL	GLDVQIKWPN	DLVVGRDKLG	GILIIETVPR
10	201	GKTVAVVIG	INFVLPEKE	NAASVQSLFQ	TASRRGNADA	AVLLETLLE
	251	LDAVILLQY	DGFAPFAEY	QAANRDHGKA	VLLLRDGETV	FEGTVKGVGD
	301	QGVLLHLETA	EKGTVVSGEI	SLRSDDRPVS	VPKRRDSERF	LLLDGNSRL
	351	KWAWVENGT	ATVGSAPYRD	LSPLGAEWAE	KADGNVRIVG	CACVCEFFKA
15	401	QVQEQRLARK	EWLPSSAQAL	GIRNHYRHPE	EHGSDRWENA	LGSRRFRSNA
	451	CVVVSCGTAV	TVDALTDGHD	YLGGTIPMGF	HLMKESLAVR	TANLNRHAGK
	501	RYFPPTTTGN	AVASGMMDAV	CGSVMMMHGR	LKEKTGAGKP	VDVLIITGGGA
	551	AKVAEALPPA	FLAENTVRVA	DNLVIYGLLN	MIAAEGREYE	HI*

Homology with the baf protein of *B. pertussis* (accession number U12020).

```

orf61 23 LLLDGGNSRLKWAUVE-NGTFATVGSAPYR----DLSPGLAEWAEEKADGNVRIVGCAVCG 77
      +L+D GNSRLK W + + A AP DL LG A R +G V G
baf 3 ILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

orf61 78 EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNLGSRFRFSRN 131
      + + L I WL + A G+RN YR+P++ G+DRW L +
baf 63 LARGEATAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGLADRWACMVGVLARQPSVHP 122

orf61 132 ACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANL 177
      +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
baf 123 PLLVASFGTATTDLTIGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167

```

35 *meningitidis:*

[illegible]

530 540 550 560 570 580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
5  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATAACGCGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
201 TTTTCGATGCC GAAGGTTTGC GCGAGCTGGG GAAAGGTTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
10  301 GCGCGGATTG CGCCGGACAA GCGGCACAAA ACCATATGTG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGCG GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTCACT TTTGGCTGGG TGTGACCG GCCGAGTAT
451 GAGTTGGGTT CGCTGTGCGC TGTGCGGCA GTGGCGTGCC GCGCGCCTT
501 GTCGCGTTTG GGTGAAAA CGCAAATCAA GTGGCCAAAC GATTGGTTCG
551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
15  601 GGCAAAACGG TTGCGGTGGT CCGTATCGGC ATCAATTTTC TGCTGCCCAA
651 GGAAGTGGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
751 CTTGATGCGG TGTGTTGCA ATATGCGCGG GACGGATTTC CGCCTTTTGT
801 GCGGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
20  851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CAAGCGGTTT TGCCTTTGGA AACGGCAGAG GGCAACAGA CGGTCGTCAG
951 CCGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTCC GTGCCGAAGC
1001 GCGGGGATTC GGAACGTTTT CTGCTGTTGG ACGCGGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGTGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
25  1101 GTACCGCGAT TGTCTGCGCT TGGGCGCGGA GTGGGCGGAA AAGTGGATG
1151 GAAATGTCCG CATCGTCGGT TGGCGCGTGT GCGGAGAATT CAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACAGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCGAGC GCGCTTCAG CCGCAACGCC
30  1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCBA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCA GCGGCATGAT
1551 GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
35  1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CCGCGGCGCG
1651 CCAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACCGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40  1  MTVLKPSHWR VLAELADGLP QHVSQARMAD MKPQQLNGF WQMPAHIRG
51  LRLQHDGYWR LVRPLAVFDA EGLRELTERS GFQTKLKEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDPRQY
151 ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGDRDLG GILLETVRTG
201 GKTAVAVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
45  251 LDAVLLQYAR DGFAPFVAEY QANRDHGA VLLLRDGETV FEGTVKGVVG
301 QGVLLHLEAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQQLARKI EWLPSAQAAL GIRNHYRHPE EHGSDFWFA LGSRRFSRNA
451 CVVVSCTAV TVDALTDGHH YLGGTIPGFI HLMKESLAVR TANLNRHAGK
50  501 RYFPFTTTGN AVASGMMDAV CGSVMMHGR LKEKTGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55  orf61a.pep  10 20 30 40 50 60
    MTVLKPSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQMPAHIRGLLRQHDGYWR
    |||||
orf61-1  MTVLKLSHWRVLAELADGLPQHVSQARMADMKPQQLNGFWQMPAHIRGLLRQHDGYWR
    10 20 30 40 50 60

60  orf61a.pep  70 80 90 100 110 120
    LVRPLAVFDAEGLRELTERS GFQTKLKECASSNDEILELARIAPDKAHKTICVTHLQSK
    |||||
orf61-1  LVRPLAVFDAEGLRELTERS GFQTKLKECASSNDEILELARIAPDKAHKTICVTHLQSK
    70 80 90 100 110 120

65  130 140 150 160 170 180

```

	orf61a.pep	GRGRQGRKWSHRLGECMLFSGFWVFDRPQYELGSLSPVAACRRALSRLGLKTQIKWPN	
	orf61-1	GRGRQGRKWSHRLGECMLFSGFWVFDRPQYELGSLSPVAACRRALSRLGLDVQIKWPN	
5		130 140 150 160 170 180	
	orf61a.pep	190 200 210 220 230 240	
	orf61-1	DLVVGRDKLGGILLETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	
10		190 200 210 220 230 240	
	orf61a.pep	250 260 270 280 290 300	
	orf61-1	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG	
15		250 260 270 280 290 300	
	orf61a.pep	310 320 330 340 350 360	
	orf61-1	QGVHLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKAWVENGTF	
20		310 320 330 340 350 360	
	orf61a.pep	370 380 390 400 410 420	
	orf61-1	ATVGSAPYRDLSPLGAEWAEEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL	
25		370 380 390 400 410 420	
	orf61a.pep	430 440 450 460 470 480	
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDGHYLGGTIMGPF	
30		430 440 450 460 470 480	
	orf61a.pep	490 500 510 520 530 540	
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP	
35		490 500 510 520 530 540	
	orf61a.pep	550 560 570 580 590	
	orf61-1	VDVITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHTX	
40		550 560 570 580 590	
	orf61a.pep	550 560 570 580 590	
45		550 560 570 580 590	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*

gonorrhoeae:

50	orf61.pep	EISLRSDXRPVSVXKRRDSERFLLLDGGNS	30
	orf61ng	TVCEGTVKGVDGRGVHLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNS	211
55	orf61.pep	RLKAWVENGTFATVGSAPYRDLSPLGAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQLAR	90
	orf61ng	RLKAWVENGTFATVGSAPYRDLSPLGAEWAEEKADGNVRIVGCAVCGESKKAQVQEQLAR	271
60	orf61.pep	KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	150
	orf61ng	KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD	331
	orf61.pep	GHYLGXGTIMGPFHLMKESLAVRTANLNRHAGKRYPFPT	189
65	orf61ng	GHYLG-GTIMGPFHLMKESLAVRTANLNRPAKRYPFPTTTGNAVASGMMDAVCGSIMM	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```

1  MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
5  51  KLGILLETIV RAGKTVAVV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLET LLAELGAVLEQ YAEFGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVLHLE TAEGEQTVVS GEISLRPDNR SVSVKPRPDS
201 ERFLLLEGGN SRLKWAVVEN GTFATVGSAP YRDLSPLGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRES RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
10 351 AVRTANLNRP AGKRYFPFTT TGNVAVSGMM DAVCGSIMMM HGRLKEKNGA
401 GKPVDVITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
451 ESEHA*

```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```

1  ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
15 51  CCGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGCGCGA TATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CCTTGCGCGT
201 TTTTCGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTCAGAGA
251 CCGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GCGCGACAAA ACCATATGCG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGTCG CACCGTTTGG
401 CCGAGTGCCT GATGTTTCA TCCGCTGGG CGTTTGACCG GCCGCGATAT
451 GAGTTGGGTT CGCTGTCGCC TGTGCGGCGA CTTGCGTGCC GCGCGCTTT
501 GGGGTGTTTG GGTTCGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
25 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
601 GGTAAAACGG TTGCGTGGT CCGTATCGGC ATCAATTTTC TGCTGCCCAA
651 GGAAGTGGA AAGCGCGCTT CCGTGCAGTC GCTGTTTCAG ACGGCATCGC
701 GCGCGGGCAA TGCCGATGCC GCGGTATTGC TGGAACATT GCTTGCGGAA
751 CTGGGCGCGG TGTGGAACA ATATGCGGAA GAAGGTTTCG CGCCATTTT
30 801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CCGCAAGGCG GTATTGCTGT
851 TGCCGCGAGG CGAAACCGTG TGCGAAGGCA CCGTTAAAGG CGTGGACGGA
901 CGAGGCGTTC TGCACTTGGA AACCGCAgaa ggcgaACAGa cggtcgtcag
951 cggcgaaaTC AGcctGCggc ccgacaacaG GTCGgtttcc gtgccgaagc
35 1001 ggcgggatTC GgaacgtTTT tTGctgttg aaggcgggaa cagccgGCTC
1051 AAGTGGGCGT GggtggAAaA cggcacgttc gcaaccgtgg gcagcgGcC
1101 gtaCCGCGAT TTGTCGCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCGGTGT GCGGAGAATC CAAAAGGCA
1201 CAAGTGAAGG AACAGCTCGC CCGAAAATC GAGTGGCTGC CGTCTTCGCG
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
40 1301 CCGTACCGTT GTTCAACGCC TTGGGCGAGC GCCGCTTTC CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
1451 AAGATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
1501 CGTTACCCCT TCCCGACCAAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
45 1551 GGACGCGGTT TGCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCGAAAGTCG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```

1  MTLVKPSHWR VLAELADGLP QHVSQALAREA DMKPQQLNGF WQMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY
55 151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILLETVRAG
201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVVDG
301 RGVHLHLETA GEQTVVSGEI SLRPDNRSVS VPKRPDSERF LLEGGNSRL
351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
401 QVKEQLARKI EWLPSAQAL GIRNHYRHPE EHGSDFWENA LGSRRFSRNA
60 451 CVVSCGTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRPAKG
501 RYFPFTTTGN AVASGMMDAV CGSIMMHGR LKEKNGAGKP VDVITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR	60
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTCVTHLQSK	120
10	orf61ng-1.pep	GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVDRPQYELGSLSPVAACRRALSRLGLDVQIKWPN	180
15	orf61ng-1.pep	DLVVGRDKLGGILITVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILITVRTGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
20	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEFGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVVDG	300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFAEYQAANRDHGKAVLLLRDGETVFEGTVKGVVDG	300
25	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSEFLLLEGGSRLKAWVWENGTF	360
	orf61-1	QGVILHLETAEGKQTVVSGEISLRSDRPVSVPKRRDSEFLLLDGGSRLKAWVWENGTF	360
30	orf61ng-1.pep	ATVGSAPYRDLSPGLAEWAEEKADGNVIRVGCACVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVIRVGCACVCGEFKKAQVQEQLARKIEWLPSSAQAL	420
35	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRRESRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRRESRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
40	orf61ng-1.pep	HLMKESLAVRTANLNRPAKRYFPPTTTGNAVASGMMDAVCGSIMMHGRLKEKNGAGKP	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYFPPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHAX	593
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMI AEGREYEHIX	593

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
50	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCGC	CCGTCATGTC
	151	GGCAAGATTC	CGCTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCGGAC	ACTTTTCTT	CAACGACAAA	GCGCTGCCT	ACCACTGGAT
55	351	ATGCGGCGCG	GCGGCATTG	CCGGTGTTCG	GCTGCTGATG	GCGGGCGGTG
	401	CGGaAGAGGG	CGCGaAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAA	GGCTGATTGC
	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
60	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGCTG	TATTGGGTT	TGGGTGC..

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 5 201 WSVGMVLSLL YLGLGC..

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATGTATG GTCGGCGTGC
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCTGCGG CCGTCATGTC
 151 GGCAAGATTG CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
 251 CCGCCGCCAG CGCATCGGTC ATTGTGCGAC TCGAGCCGCT GCTGATGGTG
 301 TTTGTGCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGGCGCG GCGGCATTTC CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
 15 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
 451 GCGGCGCGCG GCTTTTGTGC CGTATGCGT CCGACGCAA GGCTGATTGC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
 551 TGATGTGCCG CCGGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGGCG
 651 CTGTTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA
 20 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCG CGTGCTGCTG
 751 GCGGTTTTGA TTTTGGGCGA ACACCTGTGC CCCGTGTCCG CCTTGGGCGT
 801 GTTTGTCTGC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAA
 851 AATAA

25 This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 30 201 WSVGMVLSLL YLGLGCGWYA YWLWNKMSR VPANVSGLLI SLEPVVGVLL
 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

35 Orf62 1 MFYQILALI WSSSFIAAKY VYGGIDPALM VGVRLIAAL PALPACRRHV GKIPREEWKP 60
 M YQILAL+IWSS I K Y +DP L+V VR R KI + K
 HI0976 1 MLYQILALI WSSSLIVGKLTYSMDPVLVQVRLIAMIIVMPLFLRRWKIDKPMRQ 60
 Orf62 61 LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMV FVGHHFFNDK ARAYHWICGA 114
 40 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+V FVGHHFF K +
 HI0976 61 LWWLAFFNYTAVFLQFIGLKYTSAASAVTMIGLEPLL VV FVGHHFFKQNGF 114

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N.*

45 *meningitidis*:

orf62.pep 10 20 30 40 50 60
 MFYQILALI WSSSFIAAKY VYGGIDPALM VGVRLIAAL PALPACRRHV GKIPREEWKP
 orf62a MFYQILALI WSSSFIAAKY VYGGIDPALM VGVRLIAAL PALPACRRHV GKIPREEWKP
 50 10 20 30 40 50 60
 orf62.pep 70 80 90 100 110 120
 LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMV FVGHHFFNDK ARAYHWICGA
 orf62a LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMV FVGHHFFNDK ARAYHWICGA
 55 70 80 90 100 110 120
 orf62.pep 130 140 150 160 170 180
 AAFAGVALLM AGGAEEGGEV GWFGCLLVLL AGAGFCAAMR PTQRLIARIG APAFTSVSIA

-178-

```

      |||
orf62a  AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
      130      140      150      160      170      180
5
      190      200      210
orf62.pep AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
      |||
orf62a  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
      190      200      210      220      230      240
10
orf62a  SLEPVVGVLAVLILGEHLS PVSVLGVFVIAATLVAGRLSHQKX
      250      260      270      280

```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

```

15      1  ATGTTTACC  AAATCCTGCG  CCTGATTATC  TGGAGCAGCT  CGTTTATTGC
      51  CGCCAAATAT  GTCTATGGCG  GCATCGATCC  CGCATTGATG  GTCGGCGTGC
      101  GCCTGCTGAT  TGCTGCGCTG  CCTGCACTGC  CCGCCTGCCG  CCGTCATGTC
      151  GGCAAGATTC  CGCGTGAGGA  ATGGAAGCCG  TTGCTGATTG  TGTGCTTCGT
      201  CAACTATGTG  CTGACCTGCG  TACTTCAGTT  TGTGCGGTTC  AAATACACTT
      251  CCGCCGCCAG  CGCATCGGTC  ATTGTCCGAC  TCGAGCCACT  GCTGATGGTG
      301  TTTGTCGGAC  ACTTTTCTT  CAACGACAAA  GCGCGTGCCT  ACCACTGGAT
      351  ATGCGGCGCG  GCGGCATTTC  CCGGTGTCGC  GCTGCTGATG  GCGGGCGGTG
      401  CGGAAGAGGG  CGGCGAAGTC  GGCTGGTTCG  GCTGCCTGCT  GGTGTTGTTG
      451  GCGGGCGCGG  GCTTTTGTGC  CGCTATCGCT  CCGACGCAAA  GGCTGATTGC
      501  ACGCATCGGC  GCACCGGCAT  TCACATCTGT  TTCCATTGCC  GCCGCATCGT
      551  TGATGTGCCT  GCGGTTTTCG  CTTGCTTTGG  CGCAAAGTTA  TACCGTGGAC
      601  TGGAGCGTCG  GAATGGTATT  GTCGCTGCTG  TATTTGGGCG  TGGGGTGCAG
      651  CTGGTACGCC  TATTGGCTGT  GGAACAAGGG  GATGAGCCGT  GTTCTTGCCA
      701  ACGTTTCGGG  ACTGTTGATT  TCGCTCGAAC  CCGTCGTCGG  CGTGTGCTG
      751  GCGGTTTGA  TTTTGGGCGA  ACACCTGTGC  CCGGTGTCCG  TCTTGGGCGT
      801  GTTTGTCGTC  ATCGCCGCCA  CCTTGGTTGC  CGCCCGGCTG  TCGCATCAAA
      851  AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 246>:

```

35      1  MFYQILALII  WSSSFIAAKY  VYGGIDPALM  VGVRLIIAAL  PALPACRRHV
      51  GKIPREEWKP  LLIVSFVNYV  LTLLLQFVGL  KYTSAASASV  IVGLEPPLMV
      101  FVGHHFFNDK  ARAYHWICGA  AAFAGVALLM  AGGAEEGGEV  WFGCLLVLL
      151  AGAGFCAAMR  PTQRLIARIG  APAFTSVSIA  AASLMCLPFS  LALAQSYTVD
      201  WSVGMVLSLL  YLGVGCSWYA  YWLWNKGMSR  VPANVSGLLI  SLEPVVGVL
      251  AVLILGEHLS  PVSVLGVFV  IAATLVAGRL  SHQK*

```

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

```

40      orf62a.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV 60
      orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV 60
45
      orf62a.pep  LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPPLMVFGHHFFNDKARAYHWICGA 120
      orf62-1      LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPPLMVFGHHFFNDKARAYHWICGA 120
50
      orf62a.pep  AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
      orf62-1      AAFAGVALLMAGGAEEGGEVWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
55
      orf62a.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI 240
      orf62-1      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI 240
60
      orf62a.pep  SLEPVVGVLAVLILGEHLS PVSVLGVFVIAATLVAGRLSHQKX 285
      orf62-1      SLEPVVGVLAVLILGEHLS PVSALGVFVIAATLVAGRLSHQKX 285

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:

	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP	60
5	orf62.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
10	orf62.pep	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC	216
15	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
20	101	GCCGTGCTGAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCGCG	CCGTCATGTC
	151	GGCAAGATTC	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCTGT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTG	AAATACACTT
	251	CCGCGCGCGG	CGCATCGGTC	ATTGTCGGAC	TCCAGCCGCT	GCTGATGGTG
	301	TTTGTCCGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
25	351	ATGCGGCGCG	GCGGCATTTG	CCGGTGTGCG	GCTGCTGATG	GCGGGCGGGTG
	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCTGCTG	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
30	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGTTG	TATTGGGTT	TGGGTGCGG
	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTTCTGCCA
	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTCGTCGG	CGTGTGTTG
	751	GCGGTTTTGA	TTTTGGGCGA	ACATTATCG	CCCGTGCCG	CCTGGGCGT
	801	GTTTGTCTC	ATCGCCGCCA	CTTCGCCGC	CGGCCGGCTG	TCGCGCAGGG
	851	ACGCGCAAAA	CGGCAATGCC	GTCTGA		

35 This encodes a protein having amino acid sequence <SEQ ID 248>:

	1	MFYQILALII	WGSFIAAKY	VYGGIDPALM	GVRLIIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLLQLFVGL	KYTSAAASV	IVGLEPLLMV
	101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGGEV	GWFGCLLVLL
40	151	AGAGFCAAMR	PTQRLIARIG	APAFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSL	YLGLGCGWYA	YWLWNKGMSR	VPANASGLLI	SLEPVVGVL
	251	AVLILGEHLS	PVSALGVFVV	IAATFAAGRL	SRRDAQNGNA	V*

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
45	orf62ng.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHVKGIPREEWKP					
		10	20	30	40	50	60
50	orf62ng.pep	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLQLFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
55	orf62ng.pep	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
	orf62-1	AAFAGVALLMAGGAEEGGEVGVFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA					
		130	140	150	160	170	180
60	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI					
		190	200	210	220	230	240
65	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRV PANASGLLI					
		190	200	210	220	230	240

10

15

20

25

30

35

45

50

55

1	MRRFLPIAAI	CAXXLXXGLT	AATGSTSSLA	DYFWWIVAFS	AMLLLVLSAV
51	LARYVILLKK	DRRDGVFGSX	XAKXPXXMFM	TLVAXLPGVF	LFGFPQAFIN
101	GTINSWFGND	THEALERSLN	LSKSALNLA	DNALGNVAVP	QDLIDGAASL
151	PGDMGRVLEH	YAGSGFAQLA	LYNXASGKIE	KSINPHKLDQ	PFPGKARWEK
201	IQRAGSVRDL	ESIGGVLYAQ	GWFLSATHTXG	RDYALFFRQP	VPKGVAEDAV
251	LIEKARAKRA	ELSYSKKGLQ	TFLLATLLTA	SLLSIFIALV	MALYFARRFY

301 EPVLSLAEGA KAVAQGDFSQ TRPVLRNDEF GRLTXLFNHM TEQLSIAKDA
 351 DERNRRREEA ARHYLECVLE GLTTGVVVFDE EQGCLKTFNK AAGT. .

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
 5 51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101 GGTGGATTGT TGCCTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
 151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
 201 CGGTTTCGAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
 251 TACTGCCCCG CGTGTTCCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
 10 301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
 351 CCTCAATTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG
 401 GCAACGCCGT CCCCGTGCAG ATAGACCTCA TCGGCGCGGC TTCCCTGCCC
 451 GGGGATATGG GCAGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA
 501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
 15 551 CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
 601 CAACGGGCGG GTTCGGTCAG GGATTGGAA AGCATAGGCG GCGTATTGTA
 651 CGCGCAGGCG TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT
 701 TGTTTTTCG TCAGCCGGTT CCCAAAGCGG TGGCAGAGGA TGCCGTCTTA
 751 ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAAGG
 20 801 TTTCAGACCT TTTTTCCTGG CAACCCTGCT GATTGCCTCG CTGCTGTGTA
 851 TTTTCTTTCG ACTGGTCATG GCATGTATTT TCGCCCGCGG TTTCTGTGAA
 901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
 951 CAGCCAGACG CGCCCGCTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
 1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
 25 1051 GAGCGCAACC ACCGGCGCGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
 1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
 1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
 1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
 1251 GTGCTGCTGT GCCGAAGTGT TTGCGGCCAT CGGCGCGGCG GCAGGTACGG
 30 1301 ACAAAACCGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
 1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACGGCAACG GCGTGGTAAT
 1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
 1451 GGGCGGAAGT GGGCAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCAG
 1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
 35 1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
 1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG
 1651 CGTTCCCTTT CGCTCAAATT GGAAAATCAG GATTGTAACG CCTTAATCGG
 1701 CGATGTGTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
 1751 TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
 40 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
 1851 TGTGCCCGAA GTCAGGTAAT AATCGGAAAC AGGGCAGGAC GGTCCGATTG
 1901 TCCTGACGGT TTGCGACAAC GGCAGAGGGT TCGGCAGGGA AATGCTGCAC
 1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
 2001 TCTGCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
 45 2051 TGAGCAATCA GGATGCGGGT GGCCTGTGTG TCAGAAATCAT CTTGCCAAAA
 2101 ACGGTAAGAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
 50 51 LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
 101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVVPQ IDLIGAASLP
 151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
 201 QRAGSVRDLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
 251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
 301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEF GRLTKLFNHMT EQLSIAKEAD
 55 351 ERNRRREEEA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
 401 PLWGSSRHGW HGVSAQQSLL AEFVAAIGAA AGTDKPVHVK YAAPDDAKIL
 451 LGKATVLPED NGNGVVMVID DITVLIHAQK EAAGGEVAKR LAHEIRNPLT
 501 PIQLSAERLA WKLGGKLDQEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
 551 RSPSLKLENQ DLNALIGDVL ALYEAGPCR FAEELAGEPLT VAADTTAMRQ
 60 601 VLHNIFKNAA EAAEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH
 651 NAFEPYVTDK PAGTGLGLPV VKKIIIEHGG RISLSNQDAG GACVRIILPK
 701 TVKTYA*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N.*

meningitidis:

5	orf64.pep	10 20 30 40 50 60	MRRFLPIAAICAXXLXGLTAATGSTSSLADYFWWIVAFSAML LLLVLSAVLARYVILLK
	orf64a	10 20 30 40 50 60	MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAML LLLVLSAVLARYVILLK
10	orf64.pep	70 80 90 100 110 120	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN
	orf64a	70 80 90 100 110 120	DRRDGVFGSQAIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN
15	orf64.pep	130 140 150 160 170 180	LSKSALNLAADNALGNAPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE
	orf64a	130 140 150 160 170 180	LSKSALNLAADNALGNAIPVQIDXIGAASLPDGMGRVLEHYAGSGFAQLALYNAASGKIE
20	orf64.pep	190 200 210 220 230 240	KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP
	orf64a	180 190 200 210 220 230	KSINPHKLDQPFPGKARWEKIQAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP
30	orf64.pep	250 260 270 280 290 300	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
	orf64a	240 250 260 270 280 290	VPKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
35	orf64.pep	310 320 330 340 350 360	EPVLSLAEGAKAVAQGDQSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA
	orf64a	300 310 320 330 340 350	EPVLSLAEGAKAVAQGDQSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKEADERNRRREEA
40	orf64.pep	370 380 390	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT
	orf64a	360 370 380 390 400 410	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAEEQILGMPLTPLWGSSRHGWHGVSAQQSL
45	orf64a	420 430 440 450 460 470	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXXNGVVMVIDDITVLIHAQ

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

50	1	ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
	51	CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
	101	GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
	151	TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
55	201	CGGTTTCGAG ATTGCCAAAC GCCTTTCCGG GATGTTTACG CTGGTTGCCG
	251	TACTGCCCCG CGTGTTCCTG TTCGCGGTTT CCGCAGATT TATCAACGGC
	301	ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
	351	CCTCAATTTG AGCAAGTCCG CATTGAATCT GCGCGCAGAC AACGCCCTTG
60	401	GCAACGCCAT CCCCCTGCAG ATAGACNTCA TCGGCGCGGC TTCCCTGCCC
	451	NGGGATATGG GCAGGTGCT GGAACATTAC GCCGCGAGCG GTTTTGCCCA
	501	GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
	551	CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAATAATC
65	601	CAACAGGCGG GTTCGGTCAG GGATNNGGAA AGCATAGGCG GCGTATTGTA
	651	CGCGCANGGC TGGCTGTCGG CAGNNACGCA CAACGGGCGC GATTACGCCT
	701	TGTTTTTCCG TCAGCCGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
	751	ATCGAAAAGG CAAGGCGGNA ANANNNTNAG TTGAGTTACA GCAAAAAGG
	801	TTTGACAGAC TTTTTCCTNG CAACCCTGCT GATTGCCTCN CTGCTGTCGA
	851	TTTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCGCGCG TTTCGTCGAA

-183-

5
 10
 15
 20
 25

```

901 CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
951 CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGCGCGCA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGAGAGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301 ACAAAACCGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCAGC
1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAAGTGG GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCCTTT CGNCTCAATT GGAAATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGGTTT GCGGCGGAAC
1751 TTGCCGCGCA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TGTGCCCAGG GTCAGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG
1901 TCCTGACAGT TTGCACAAC GGAAGGGGT TCGGAGGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001 ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGCG CNCATCAGCC
2051 TGAGCAATCA GGATCGGGG GCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
  
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

30
 35
 40

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFNG
101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDXE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARFEVE
301 PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLCVLEG LTTGVVVFDE QGCLTKFNKA AEQILGMPLT
401 PLWGSRRHGW HGVSQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551 RSPSQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601 VLHNIKFNAE EAAEEADVPE VRVKSEAGQD GRIVLTVCDN KGKFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIIIEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*
  
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

45
 50
 55
 60
 65

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              |||
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLDYFWWIVAFSAML VLSAVLARYVILLK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              |||
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90      100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              |||
orf64-1      SKSALNLAADNALGNAIPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQFPFGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
              |||
orf64-1      SINPHKLDQFPFGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
              190     200     210     220     230     240

              250     260     270     280     290     300
  
```


-184-

	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKKGLOTFFLATLLIASLLSIFLALVMALYFARRFVE	
	orf64-1		
5		250 260 270 280 290 300	
	orf64a.pep	310 320 330 340 350 360	
	orf64-1		
10		310 320 330 340 350 360	
	orf64a.pep	370 380 390 400 410 420	
	orf64-1		
15		370 380 390 400 410 420	
	orf64a.pep	430 440 450 460 470 480	
	orf64-1		
20		430 440 450 460 470 480	
	orf64a.pep	490 500 510 520 530 540	
	orf64-1		
25		490 500 510 520 530 540	
	orf64a.pep	550 560 570 580 590 600	
	orf64-1		
30		550 560 570 580 590 600	
	orf64a.pep	610 620 630 640 650 660	
	orf64-1		
35		610 620 630 640 650 660	
	orf64a.pep	670 680 690 700	
	orf64-1		
40		670 680 690 700	
	orf64a.pep	PAGTGLXLPVVKIIEEHGGXISLSNQDAGGAXVRIILPKTVETYAX	
	orf64-1		
45		670 680 690 700	

Homology with a predicted ORF from *N.gonorrhoeae*ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.**gonorrhoeae*:

	orf64.pep	MRRFLPIAIAICAXXLXXGLTAATGSTSSLDYFWWIVAFSAMPLLLVLSAVLARYVILLK	60
	orf64ng		
	orf64ng	MRRFLPIAIAICAVVLLYGLTAATGSTSSLDYFWWIVSFSAMPLLLVLSAVLARYVILLK	60
55		DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	120
	orf64ng		
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
60		LSKSALNLAADNALGNAPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng		
	orf64ng	LSKSALDLAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNASGKIE	179
	orf64.pep	KSINPHKLDQFPFGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
	orf64ng		
65		KSINPHQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWIIVSFS AMLLLVLSAV
15 51 LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGFLF FGLSAQFNG
101 TNSWFNGND HEALERSLNL SKSALDLAAD NAVSNVAVPQ IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQDFQP LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT FFLVTLIIAS LLSIFIALVM ALYFARRFVE
20 301 PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNMHT EQLSIAKEAD
351 ERNRRREEAA RAYHECVLDG LTTGVVVSYP LSCCRTAVFS TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>

1 ATGCGCCGCT TCCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGCTGTA
25 51 CGGATTGACG GCGGCGACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
101 GGTGGATAGT CTCGTTACAG GCAATGCTGC TGCTGTGTGT GTCCCGCGTT
151 TTGGCACGTT TACGTCATATT GCTGTTGAAA GACAGGCGCA ACCGCGTGTT
201 CGGTTCCGAG ATTGCCAAAC GCCTTCCGG GATGTTACAG CTGGTCGCCG
251 TACTGCCCGG CTTGTTCTCG TTCGCGCATT CCGCGCAGTT TATCAACGGC
30 301 ACGATTAATT CGTGTTTCGG CAACGACACC CCGGACGCC TCGAACGCGAG
351 CCTTAATTG AGCAAGTCCG CACTGGATTG GCGGCGAGAC AATGCCGTCA
401 GCAACGCCGT TCCCGTACAG ATAGACCTCA TCGGCACCGC CTCCTGTGCG
451 GGCAATATGG GCAGTGTGCT GGAACACTAC GCGCGCAGCG GTTTTGCCCA
501 GCTTGCCCTG TACAATGCCG CAAGCGGGAA AATCGAAAAA AGCATCAATC
35 551 CGCACCAATT CGACCAGCCG CTTCCCGACA AAGAACATTG GGAACAGATT
601 CAGCAGACCG GTTCGGTTCG GAGTTTGGA AGCATAGGCG GCGTATTGTA
651 CGCGCAGGGA TGGTTGTCGG CAGGTACGCA CAACGGGCGC GATTACGCGC
701 TGTCTCTCCG CCAGCCGATT CCCGAAAATG TGCGACAGGA TGCCGTCTCTG
751 ATTGAAAAGG CGCGGGCGAA ATATGCCGAA TTGAGTTACA GCAAAAAAGG
40 801 TTTGCAGACG TTTTCTCTG TAACCTGTCT GATTGCCTCA CTGCTGTGCA
851 TTTTCTCTGC GCTCGTAAAT GCACGTGATT TGCCCGCGCG TTTCTGCGAA
901 CCCATTCTGT CGCTTGCCGA GGGCGCAAAG CCGGTGGCGC AGGGTGATTT
951 CAGCCAGACG CGCCCCGTAT TGCGCCAACA CGAGTTCCGA CGTTTGACCA
1001 AGCTGTTCAA CCATATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
45 1051 GAACGCAACC GCCGCGCGCA GGAAGCGGCC CGTCACTACC TCGAGTGCCT
1101 GTTGGATGGG TTGACTACCG GTGTGGTGGT GTTTGACGAA AAAGGCCGTT
1151 TGAAAACCTT CAACAAGGCG GCGGAACAGA TTTTGGGGAT GCGCGTCGCC
1201 CCCTCTGGG GCAGCGCCG GCACGTTGG CACCGCGTTT CCGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TtgccgccAT CGGTGCGGCG GCAGGTACGG
50 1301 ACAAAACCGT CCAGGTGGA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGC CGACGGTATT GCGCGAAGAC AACGGCAAC GCGTGGTGAT
1401 GGTGATTGAC GACATCACCG TGCTGATACG CGCGCAAAAA GAAGCCGCGT
1451 GGGGTGAAGT GGGCAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCAGC
1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAATTTGG GCGGGAAGCT
55 1551 GGACGATCAG GACGCGCAAA TCCTGACGCG TtcgACCGAC ACCATCATCA
1601 AACAGgtggc gGCGTTAAAA GAAATGGTCG AGGCATTCCG CAATTACGCG
1651 CGCGCCCCCT CGCTCAAAC GGAATATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTTTTG GCCCTGTACG AAGCCGGCCC GTGCCGTTT GAGCGGGAAC
1751 TTTGCCGCGA ACCCGTGTAT ATGGCGCGG ATACGACCGC CATTGCGGAG
60 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGCGAGGAC GGACGGATTG
1901 TCCTGACGGT TTGCGACAA GGCAAGGGAT TCGGCAAGGA AATGCTGCAC
1951 AATGCTTTTC AGCCGTATGT GACGGATAAG CCGGCGGGAA CGGGACTGGG
2001 TCTGCCTGTA GTGAAAAAAA TCATTGGAGA ACACGGCGCG CGCATCAGCC
2051 TGAGCAATCA GGATGCGGGT GGGGCGTGTG TCAGAATCAT CTTGCCAAAA
65 2101 ACGGTAGAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

```

1  MRRELPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
51  LARYVILLK  DRRNGVFGSQ  IAKRLSGMFT  LVAVLPGLFL  FGISAQFING
101 TINSWFGNDT HEALERSLNL  SKSALDLAAD  NAVSNAVFPV  IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL  YNAASGKIEK  SINPHQFDQP  LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQG  WLSAGTHNGR  DYALFFRQPI  PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT  FFLVTLLIAS  LLSIFLALVM  ALYFARRFVE
301 PILSLAEGAK AVAQGDFSQT  RPVLRNDEFG  RLTKLFNHMT  EQLSIAKEAD
351 ERNRRREEAA RHYLECVLDG  LTTGVVVFDE  KGRKTFNKA  AEQILGMPLA
401 PLWGSSRHGW HGVSAQQSLL  AEVFAAIGAA  AGTDKPVQVE  YAAPDDAKIL
451 LGKATVLPED NGNGVVMVID  DITVLIRAQK  EAAWGEVAKR  LAHEIRNPLT
501 PIQLSAERLA WKLGGKDDQ  DAQILTRSTD  TTIKQVAALK  EMVEAFRNYA
551 RAPSLKLENQ DLNALIGDVL  ALYEAGPCRF  EAELAGEPLM  MAAOTAMRQ
601 VLHNIKNAE  EAAEEADMPE  VVRKSETGQD  GRIVLTVCDN  GKGFGKEMLH
651 NAFEPYVTDK PAGTGLGLPV  VKKIIGEHHG  RISLSNQDAG  GACVRIILPK
701 TVETYA*

```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

```

20  orf64ng-1.pep  10      20      30      40      50      60
    MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVSFSAMLLLVLSAVLARYVILLK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFSAMLLLVLSAVLARYVILLK
    10      20      30      40      50      60

25  orf64ng-1.pep  70      80      90      100     110     120
    DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
    70      80      90      100     110     120

30  orf64ng-1.pep  130     140     150     160     170     180
    SKSALDLAADNAVSNVFPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  SKSALNLAADNALGNVFPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
    130     140     150     160     170     180

35  orf64ng-1.pep  190     200     210     220     230     240
    SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQPI
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  SINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
    190     200     210     220     230     240

40  orf64ng-1.pep  250     260     270     280     290     300
    PENVAQDAVLIEKARAKYAE LSYSKKGLQTFFLVTLLIASLLSIFLALVMALYFARRFVE
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  PKGVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
    250     260     270     280     290     300

45  orf64ng-1.pep  310     320     330     340     350     360
    PILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
    310     320     330     340     350     360

50  orf64ng-1.pep  370     380     390     400     410     420
    RHYLECVLDGLTTGVVVFDEKGRKTFNKAAEQILGMPLAPLWGSSRHGW HGVSAQQSLL
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  RHYLECVLEGLTTGVVVFDEQCLKTFNKAAEQILGMPLTPWGSSRHGW HGVSAQQSLL
    370     380     390     400     410     420

55  orf64ng-1.pep  430     440     450     460     470     480
    AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK
    430     440     450     460     470     480

60  orf64ng-1.pep  490     500     510     520     530     540
    EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKDDQDAQILTRSTD TTIKQVAALK
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf64-1  EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKDDQDAQILTRSTD TTIKQVAALK
    490     500     510     520     530     540

```

-187-

5	orf64-1	EAANGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDEQDAQILTRSTDITIVKQVAALK	490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLMMAADTTAMRQ	550	560	570	580	590	600
10	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLTVAADTTAMRQ	550	560	570	580	590	600
	orf64ng-1.pep	VLHNIKFNAEAAEEADMPVVRVKSETQDGRIVLTVCNKGKFGKEMLHNAFEPYVTDK	610	620	630	640	650	660
15	orf64-1	VLHNIKFNAEAAEEADVPVVRVKSETQDGRIVLTVCNKGKFGREMLHNAFEPYVTDK	610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRIISLNSQDAGGACVRILPKTVETAYX	670	680	690	700		
20	orf64-1	PAGTGLGLPVVKKIIEHGGRIISLNSQDAGGACVRILPKTVKTYAX	670	680	690	700		

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

25	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir S18624 ntry protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product {Azorhizobium caulinodans} Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)							
	Query: 7	IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXXRYVILLKDRRNGV	66					
30		I+A+ ++L GLT + + + R + + K R G						
	Sbjct: 35	ISALATFLILMGLTPVVPVTHQVVIS----VLLVNAAVLILSAMVGREIWRIAKARAGR	90					
35	Query: 67	FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD	126					
		+++ R+ G+F +V+V+P + + +++ ++ ++ WF T E + S+++++ +						
40	Sbjct: 91	AAARLHRIIVGLFAVVSVPVAILVAVVASLTLDRLDRWFSMRTQEIIVASSVSVAQTYVR	150					
	Query: 127	LAADNAVSNAVVPQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP	184					
45		A N + + + DL S+ Y G S F Q+ AA + ++						
	Sbjct: 151	EHALNIRGDILAMSADLTRKSV-----YEGDRSRFNQILTAQAALRNLPGLMI	200					
50	Query: 185	HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA-----	233					
		+ D + ++ + I + V + +IG Q + N DY						
55	Sbjct: 201	RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQFVIYLP--NDADYVAAVPLKDYDD	256					
	Query: 234	--LFFRQPIPENVAQDAVLIEKARAKYAELSYSKGLQTFVLVTXXXXXXXXXXXXXVMA	291					
60		L+ + I V ++ A Y L + G+Q F + +						
	Sbjct: 257	LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG	316					
65	Query: 292	LYFARRFVEPILSLAEGAKAVAQGDFTSRPVLRLND-EFGRLTKLFNMHTEQLSIXXXXX	350					
		L F++ V PI L A VA+G+ P+ R + + L + FN MT +L						
70	Sbjct: 317	LNFSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI	376					
	Query: 351	XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRKLTFNKAAEQILGMPLAPLWGSSRHGW	410					
75		+ E VL G+ GV+ D + R+ N++AE++LG L+ + RH						
	Sbjct: 377	LTARDQIDSRRTFEAVLSGVGAGVIGLDSQERITILNRSARLLG--LSEVEALHRLHA	434					
80	Query: 411	HGVSAQQSLLAEVFXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM	467					
		V LL E + VQ D + + V E + +G V+						
85	Sbjct: 435	EVVPETAGLLEEA-----EHARQRSVQGNITLTRDGRERVFAVRVTTEQSPEAEHGWV	488					
	Query: 468	VIDDITVLIRAQKEAANGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDDQDAQILTR	527					
90		+DDIT LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL K G + QD +I +						
	Sbjct: 489	TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ	547					
95	Query: 528	STDITIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE	587					
		TDTII+QV + MV+ F ++AR P +++QD++ +I + L G +						
100	Sbjct: 548	CTDTIIRQVGDIGRMVDEFSSFARMFKPVVDSQDMSEIIRQTVFLMRVGHPEVVFDESEVP	607					
	Query: 588	PLMMAA-DTTAMRQVLHNIKFKNXXXXXXXXXDMPEVRVK-----SETQDGRIVLTVCN	639					
105		P M A D + Q L NI KN P+VR + + G+D +V+ + D						
	Sbjct: 608	PAMPARFDRRLVSQALTNILKNAAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID	664					

Query: 640 NGKGFGEKMLHNAFEPYVTDKPAAGTGLGLPVVKKIIGEHGGRISLSNQDAG-GACVRIIL 698
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L
 Sbjct: 665 NGTGLPQESRNRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACCT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTG AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGC
451 CACGCGTTGG ATACG...

```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```

1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPNASTVIG
151 HALDT...

```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCGGCATC CACACCACCT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTGTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTG AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
451 AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCTTCTT ACGCAAGCAG
501 CGATGGATTG ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```

1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGIAT VDYLFLKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

-189-

orf66 1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
 M F+ Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
 o221 1 MNVFSQTRYKALEWLSLFLHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

5 orf66 61 RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
 RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
 o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

10 orf66 121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
 +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
 o221 121 LGQILDVHVFNRLRQSRRWLAPTASTLFGNVSDT 155

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf66.pep		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV					
orf66a		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV					
20		10	20	30	40	50	60
		70	80	90	100	110	120
orf66.pep		RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
orf66a		RIFGSHLARRIIFWVMFPALLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
25		70	80	90	100	110	120
		130	140	150			
orf66.pep		IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT					
		:					
orf66a		LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
30		130	140	150	160	170	180
		190	200	210	220		
orf66a		VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
35		190	200	210	220		

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

	1	ATGTACGCAT	TTACCGCCGC	ACAGCAACAG	AAGGCACTCT	TCTGGCTGGT
	51	GCTTTTTCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAGTTCC
40	101	CCTTCCAAAT	TTCCGGCATC	CACACCACTT	GGGGCGCGTT	TTCTTTTCCC
	151	TTCATCTTCC	TCGCCACCGA	CCTGACCGTC	CGCATTTTCG	GTTGCACTT
	201	GGCACGCGCG	ATTATCTTTT	GGGTCATGTT	CCCCGCCCTT	TTGCTTTCCT
	251	ACGTCTTTTC	CGTTTGTTC	CACAACGGCA	GTTGGACGGG	CTTGGCGCG
	301	CTGTCCGAAT	TCAACACCTT	TGTCGGACGC	ATCGCGCTGG	CAAGTTTTC
45	351	CGCCTACGCG	CTCGGACAAA	TCCTTGATAT	TTTGTGTTC	AACAAATTAC
	401	GCCGTCTGAA	AGCGTGGTGG	GTTGCCCGCA	CTGCATCAAC	CGTCATCGGC
	451	AACGCCCTAG	ATACGTTGGT	ATTTTCGCC	GTTGCCTTCT	ACGCAAGCAG
	501	CGATGGATTT	ATGGCGGCAA	ACTGGCAGGG	CATCGCTTTT	GTCGATTACC
	551	TGTTCAAAC	CACCGTCTGC	GGTCTGTTT	TCCTGCCCGC	CTACGGCGTG
50	601	ATTCTGAATC	TGCTGACGAA	AAAAGTACG	ACCCTGCAAA	CCAAACAGGC
	651	GCAAGACCGC	CCCGCGCCT	CGTGCAAAA	TCCGTAA	

This encodes a protein having amino acid sequence <SEQ ID 264>:

	1	MYAFTAAQQQ	KALFWLVLFH	ILIIAASNYL	VQFPFQISGI	HTTWGAFSFP
	51	FIFLATDLTV	RIFGSHLARR	IFWVMFPAL	LLSYVFSVLF	HNGSWTGLGA
55	101	LSEFNTFVGR	IALASFAAYA	LGQILDIFVF	NKLRLKAWW	VAPTASTVIG
	151	NALDTLVFFA	VAFYASSDGF	MAANWQGI AF	VDYLEFKLTV	GLFFLPAYGV
	201	ILNLLTKKLT	TLQTKQAQDR	PAPSLQNP*		

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

		10	20	30	40	50	60
orf66a.pep		MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV					
orf66-1		MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV					
60							

-190-

		10	20	30	40	50	60
5	orf66a.pep	70	80	90	100	110	120
	orf66-1	70	80	90	100	110	120
10	orf66a.pep	130	140	150	160	170	180
	orf66-1	130	140	150	160	170	180
15	orf66a.pep	190	200	210	220	229	
	orf66-1	190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae*:

25	orf66.pep	MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFI FLATDLTV	60
	orf66ng	MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFI FLATDLTV	60
30	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
	orf66.pep	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

35 The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```

1  ATGTACGCAT  TGACCGCCGC  ACAGCAACAG  AAGGCACTCT  TCCGGCTGGT
51  GCTTTTCCAT  ATCCTCATCA  TCGCCGCCAG  CAACTATCTG  GTGCAGTTC
101 CCTTCGGGAT  TTTCGGCATC  CACACCACTT  GGGGCGCGTT  TTCCTTTCCC
151 TTCATCTTCC  TCGCCACCGA  CCGTACCGTC  CGCATTTCG   GTTCGCACTT
201 GCGCGGCGG  ATTATCTTTT  GGGTGATGTT  CCCC GCCCTT  ttgCTTcat
251 aCGTCTTTT  CGTTTGTGTC  CACAACGGCA  GTTGGACGGG  CTGGGCGCG
301 ctgTCCCAAT  TCAACACCTT  TGTCCGACGC  ATCGCGCTGG  CAAGTTTTC
351 CGCCTACGCG  CTCGGACAAA  TCCTTGATAT  TTTCGTATTC  GACAAATTAC
401 GCCGTCTGAA  AGCGTGGTGG  ATTGCCCGG  CCGCATCAAC  CGTCATCGGC
451 AATGCACTGG  ACACGTTAGT  ATTTTGTGCC  GTTGCCTTTT  ACGCAAGCAG
501 CGATGAATTT  ATGGCGGCAA  ACTGGCAGGG  CATCGCTTTT  GTCGATTACC
551 TGTTCAAAC  TACCGTCTGC  ACCCTCTTCT  TCCTGCCCGC  CTACGGCGTG
601 ATACTGAATC  TGCTGACGAA  AAAACTGACG  GCCCTGCAAA  CCAAACAGGC
651 GCAAGACCGC  CCCGTGCCCT  CGCTGCAAAA  TCCGTAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 266>:

```

1  MYALTAQQQ  KALFRLVLFH  ILIIAASNYL  VQFPFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LLSYVFSVLF  HNGSWTGLGA
101  PSQNTFVGR  IALASFAAYA  LGQILDIFVF  DKLRLKAWW  IAPAASTVIG
151  NALDTLVFFA  VAFYASSDEF  MAANWQGI AF  VDYL FKLTV C  TLFFLPAYGV
201  ILNLLTKKLT  ALQTKQAQDR  PVP SLQNP*

```

An alternative annotated sequence is:

```

1  MYALTAQQQ  KALFRLVLFH  ILIIAASNYL  VQFPFRIFGI  HTTWGAFSFP
51  FIFLATDLTV  RIFGSHLARR  IIFWVMFPAL  LLSYVFSVLF  HNGSWTGLGA
101  LSQNTFVGR  IALASFAAYA  LGQILDIFVF  DKLRLKAWW  IAPAASTVIG
151  NALDTLVFFA  VAFYASSDEF  MAANWQGI AF  VDYL FKLTV C  TLFFLPAYGV
201  ILNLLTKKLT  ALQTKQAQDR  PVP SLQNP*

```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

5  orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV  60
    orf66ng    MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV  60

10  orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
    orf66ng    RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120

15  orf66-1.pep  IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
    orf66ng    LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180

20  orf66-1.pep  VDYLFLKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX  229
    orf66ng    VDYLFLKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRVPVPSLQNPX  229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

20  sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
    REGION (O221)
>gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
(U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
hypothetical 25.3 kD protein in ftsY-nika intergenic region [Escherichia coli]
Length = 221
Score = 273 bits (692), Expect = 5e-73
25  Identities = 132/203 (65%), Positives = 155/203 (76%)

Query: 1 MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV  60
      M + Q+ KALF L LFH+L+I +SNYLVQ P I G HTTWGAFSFPFIFLATDLTV
30  Sbjct: 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV  60

Query: 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
      RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
35  Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120

Query: 121 LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
      LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
40  Sbjct: 121 LGQILDVHVFNRRLRQSRWWLAPTASTLFGNVSDTLAFFFFIAFWRSPDAFMAEHWMEIAL 180

Query: 181 VDYLFLKLTVCTLFFLPAYGVILN 203
      VDY FK+ + +FFLP YGV+LN
40  Sbjct: 181 VDYCFKVLISIVFFFLPMYGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```

50  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
    51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAyGCA GTmwrAATAT
    101 CTGAAACTGT TTCAGTTGAT ACCGACAAG GTGCGAAAAT TCATAAGTTT
    151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTAATAA AAACGGTAGA
    201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
    251 CCGCCAGCGT ATCCGCGGCC GCGTATTGG CGGGGGTCGG CAACTTGCC
    301 CGCTTAGCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
    351 CcTTTtagcc CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
    401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGTGTAAG AGGCTACGAA
    451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA

```


-192-

501 TGGCTGCTAC GCGTTGAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF
 51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLA
 5 101 RLGAKEFSTRA VPHYVTALLA HDVYETFKED IQARGYQYDP ETDKFKVKGYE
 151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
 10 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
 251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
 15 351 CCTTTAGGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
 451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
 20 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLA
 101 RLGAKEFSTRA VPHYVTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
 151 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf72.pep		MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF VPKNSKTYSS					
30 orf72a		MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf72.pep		DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLA RLGAKEFSTRA VPHYVTALLA					
35 orf72a		DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLA RLGAKEFSTRA VPHYVTALLA					
		70	80	90	100	110	120
		130	140	150	160	170	
40 orf72.pep		HDVYETFKEDI QARGYQYDP ETDKFKVKGYE YSNCLWYEDK RRINRTYGCY GVD					
orf72a		HDVYETFKEDI QARGYQYDP ETDKFAKVS GX					
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

45 1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
 101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
 151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
 50 201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
 251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCGG CAAACTTGCC
 301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
 351 CCTTTAGGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
 401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
 451 TAA

55 This encodes a protein having amino acid sequence <SEQ ID 272>:

-193-

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPHYGTALLA HDVYETFKEI IQARGYQYDP ETDKFAKVS
151 *

```

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

10 orf72a.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||||
orf72-1      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS

15 orf72a.pep      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPHYGTALLA
    |||||
orf72-1      70      80      90     100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPHYGTALLA

20 orf72a.pep      130     140     150
    HDVYETFKEI IQARGYQYDP ETDKFAKVS GX
    |||||
orf72-1      130     140     150
    HDVYETFKEI IQARGYQYDP ETDKFAKVS GX

```

Homology with a predicted ORF from *N. gonorrhoeae*

25 ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae*:

```

30 orf72.pep      MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS 60
    || : |||||
orf72ng      MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF VPKSSNIYSS 60

orf72.pep      DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPHYGTALLA 120
    || : |||||
orf72ng      DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LVRQAKFGT RA VPHYGTALLA 120

35 orf72.pep      HDVYETFKEI IQARGYQYDP ETDKFKVGYEY SNCLWYED KRRINR TYGCGYVD 173
    ||||| : |||||
orf72ng      HDVYETFKEI IQARGCRYDP ETDKFKVGYEY ANCLWYED RRRINR TYGCGYVD SSIMRLM 180

```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

40 1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
    51 VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLV
    101 RQGAKEGTRA VPHYGTALLA HDVYETFKEI IQARGCRYDP ETDKFKVGYE
    151 YANCLWYEDE RRINR TYGCGY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
45 201 ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGGCAVN KGDDFRAGAS
    251 FSLGRNPKYK EEMDAKKPEE ILSLKVDADP DKYIEATGYP GYSEKVEVAP
    301 GTKVNMGPVT DRNGNPVQVA ATFGDRDAQN TTADVQVIPR PDLTPASAEA
    351 PHAQPLPEVS PAENPANNDP PDENPGTRPN PEPDPLNDP ANPDTGQPG
    401 TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
    451 HDISIPQVTD DKTWSSHNF LPSNGVCPQPK TFHVFGQYR ASYEPLCVEA
50 501 EKIRFAVLLA FIIMSAFVVF GSLGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

55 1  ATGGTCACAA AACATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
    51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
    101 CTGAAACTCT TTCGGTTGAT ACCGACAAG GCGCGAAAGT TCATAAGTTC
    151 GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTTAACAA AAGCGGTAGA
    201 TTTAACGCAT ATCCCCACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
    251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGT CGGGGGTCGG CAAACTTGTC
    301 CGCCAAGGCG CGAAATTCGG CACAAGGCG GTTCCCTATG TCGGAACAGC
    351 CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
60 401 GAGGCTGCCG ATACGATCCC GAAACCGACA AATTT

```

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

```

1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAHVHKF
51 VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGKLV
101 RQGAKEFGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKF

```

5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```

10 orf72ng-1.pe 10 20 30 40 50 60
    || :|||||:|||||:|||||:|||||:|: |||
orf72-1        10 20 30 40 50 60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSDT GQGAKEHVKF VPKNSKTYSS

15 orf72ng-1.pe 70 80 90 100 110 120
    || :|||||:|||||:|||||:|: |||:|||||:
orf72-1        70 80 90 100 110 120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLGVGKLAR LGAKFSTR AVPYVGTALLA

20 orf72ng-1.pe 130 140
    HDVYETFKEDI QARGCRYDP ETDKF
orf72-1        130 140 150
    HDVYETFKEDI QARGYQYDP ETDKFAKVSGX

```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

```

30 1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
    51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
    101 TGATGGCGGC AGGTTTGGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
    151 CTGACCGGTT CTTTATTGGC CGGCGCGGCG AATGAGAAGC GCGGGAAGG
    201 TATCCGTTTA TCAGATGTTG TGGCCTATC..

```

35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

```

1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLQTG
51 LTGLLLAGAA MRSGGKVSIV QMLWPI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

```

40 1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAGATTAT
    51 GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
    101 TGATGGCGGC AGGTTTGGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
    151 CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGAGGGT
    201 ATCCGTTTAT CAGATGTTGT GGCCTATCCG TTATACGGTG GCGGCTGTGT
    251 GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTTGGCGGT ATTGCTGCTG
45 301 CTGCCGTTTA AGGGAGGGGC AGTGTTCAG GCAGGAGGTG CGGAAATTT
    351 TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
    401 ATATTATCGA GGGAGAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
    451 TCCCGAAACG CCATCGAACA CAAAAAGAC GAATAA

```

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

```

50 1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMRLHTG
    51 LSGLLLAGAA MRSGGRVSIV QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101 LPFKGGAFLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEYPGGNR

```

151 SRNAIEHKKD E*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.*

5 *meningitidis*:

```

      10      20      30      40      50      60
orf73.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLQTGLTGLLLAGAA
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73a    MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVVMRLRHTGLSGLLLAGAA
      10      20      30      40      50      60

          70
orf73.pep MRSGBKVSQYQMLWPI
          |||:|||||:|||||
orf73a    MRSGBRVSVYXMLWYIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
15

```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```

      1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTT TGGAGATTAT
    51  GTCGATTGTG TGGGTGCGG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
101  TAATGGCGGC AACCTTTGCC GCCGGCGTGG TGATGCTCAG GCATACGGGG
201  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201  ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACGGTG GCGGCGGTGT
251  GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNGGCGGT ATTGCTGNTG
301  CTNCCGTTTA AGGGAGGTGC AGTGTGTCAG GCAGGAGGTG CGGAAAATTT
351  TTTCAACATG AACCANTCGG GCAGAAAAGA NGGCNTTTC CGCGATGACG
25  401  ATATTATCGA GGGGGAATAT ACGGTGAAG ANCCTACGG CGGCANTCGT
451  TTCCGAAACG CCNTNGAACA CAAAAAGAC GAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```

      1  MRFFGIGFLV LLFLEIMSIW VADWLGGGW TLFLMAATFA AGVVMRLRHTG
    51  LSGLLLAGAA MRSGBRVSVY XMLWYIRYTV AAVCXMSPGF VSSVXAVLLX
30  101  LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151  FRNAXEHKKD E*

```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```

      10      20      30      40      50      60
orf73a.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVVMRLRHTGLSGLLLAGAA
35  orf73-1  MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLRHTGLSGLLLAGAA
          10      20      30      40      50      60

          70      80      90      100     110     120
orf73a.pep MRSGBRVSVYXMLWYIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73-1    MRSGBRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLPFKGGAVLQAGGAENFFNM
          70      80      90      100     110     120

          130     140     150     160
orf73a.pep NXSGRKXGXS RDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf73-1    NQSGRKEGFSRDDDIIEGEYTVVEEPYGGNRSRNAIEHKKDEX
50  130     140     150     160

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N.*

gonorrhoeae:

```

55  orf73.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLQTGLTGLLLAGAA 60
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

    orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLHTGLSGLLLAGAA    60
    orf73.pep    MRSGGKVSQYQMLWPI                                                    76
    orf73ng      VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM    120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

    1  ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
    51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG AcgctGTTC
    101 TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
    151 CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
    201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
    251 GTCTGatgag tCcgGATTC GTATCCTccg tggtggCGGT ATTGCTGCTG
    301 CTGCcgttta aggGaggGgc agtggtgcag gcaggaggtg cggaaaATTT
    351 TTTCACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
    401 atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
    451 tcccgAAcgc ccatcgaaca cgaAAagac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

    1  MRFFGIGFLV LLFLEIMSIW VADWLGGGW TLFLMAATFA AGVLMRLHTG
    51  LSGLLLAGAA VKSSGKVSQY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
    101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGNR
    151 SRNAIEHKD E*

```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

    10      20      30      40      50      60
    orf73-1.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAAGFAAGVLMRLHTGLSGLLLAGAA
    25      orf73ng MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLMAATFAAGVLMRLHTGLSGLLLAGAA
    10      20      30      40      50      60
    70      80      90      100     110     120
    orf73-1.pep MRSGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
    30      orf73ng VKSSGKVSQYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
    70      80      90      100     110     120
    130     140     150     160
    orf73-1.pep NQSGRKEGFSRDDDIIEGEYTVVEPYGGNRSRNAIEHKKDEX
    35      orf73ng NQSGRKEGFFHDDDIIEGEYTVVEKPDGGNRSRNAIEHKDEX
    130     140     150     160

```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

    1  ATGTTTGTTC TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
    51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
    101 TCGGCAATTT GCGCGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
    151 GCG..... GCGA AGACACGCGC GTTACCGCAC AGCTTTTGTG
    201 CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
    251 GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
    301 GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
    351 ACTCGCCCGC CGCGTGGCTG AGGCCGGGTT TAAAGTCGTT CCGTCGTGG
    401 GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
    451 GATTTTATT TCAACGGTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

```

-197-

501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
 551 CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
 601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
 701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
 751 AAACACGAAG GCTTGTCGCA GTCCGCGCAA AACATCATGA AAATCCTCAC
 801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG
 851 GCGAGGGAAA GAAAGCTTTG TACGAT..

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

10 1 MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
 51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV
 101 AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS
 151 DFYFNGFVPP KSGERRKLFA KVVRAAFPIV MFETPHRIGA ALADMAELFP
 201 ERRMLLAREI TKTFFETFLS TVGEIQTALS ADGQSRGEM VLVLYPAQDE
 15 251 KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD..

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
 151 CCGCTTACCG CACAGCTTTT GAGCGGTAC GGCATTACAG GCAAACCTCGT
 201 CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAG ATTGTGCGCT
 251 ATCTTTTCTG GCGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTTCACGG TTTGTACCG
 451 CCGAAATCGG GAGAACGAG GAAACTGTT GCCAAATGGG TGCGGGCGGC
 501 GTTCTCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTC
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG
 701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
 801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGAAAAAAC AAATAG

35 This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKVVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE
 40 201 ITKTFFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N.*

45 *meningitidis*:

10 20 30 40 50 60
 orf75.pep MFVFQTAFXMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKAXXXXAEDTR
 orf75a MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTR
 10 20 30 40 50
 70 80 90 100 110 120
 orf75.pep VTAQLLSAYGIQKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
 orf75a VTAQLLSAYGIQKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
 60 70 80 90 100 110
 130 140 150 160 170 180
 orf75.pep RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKVVRAAFPIV

20	1	ATGTTTCAGA	AACATTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCGGAC	ATTACCCGTG
	101	CGCCTTTFGC	GGCATTTGCA	AAGGCGGACA	TCATCTGTGC	CGAAGACACG
	151	CGCGTTTACG	CGCAGCTTTT	GAGCGCGTAC	GGCATTCAGG	GCAAACTCGT
25	201	CAGCGTGCGC	GAACACAACG	AACGGCAGAT	GGCGGACAA	ATTGTTCGGT
	251	ATCTTTTACA	CGGCATGGT	GTGGCACAGG	TTTCCGATGC	GGGTACGCCG
	301	CGCGTGTGCG	ACCGGGGCGC	GAAACTCGCC	CGCCGGGTGC	GTGAGGTGCG
	351	GTTTAAAGTT	GTCCCTGTTG	TCGGCGCAAG	CGCGGTGATG	GCGGCTTTGA
30	401	GTGTGGCTGG	TGTGGCGGGA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGCTTGC
	501	GTTTCCCGTG	GTGATGTTTG	AAACGCCGCA	CCGCATCGGG	GCGACGCTTG
	551	CCGATATGGC	GGAAGTGTTC	CCCGAACGCC	GATTAATGCT	GGCGCGCGAA
35	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAAG	GGCACGGTTG	GGGAATTTCA
	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTFTG
	701	TGCTTTATCC	GCGCGAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCCGCG
	751	CAAAACATCA	TGAAAATCCT	CACAGCGAG	CTGCCGACCA	AACAGGCGGC
35	801	GGAGCTTGCG	GCCAAATCA	CGGGCGAGGG	AAAAAAGCT	TTGTACGATC
	851	TGGCACTGTG	TGGAAAAAAC	AAATGA		

40

1	MFQKHLQKAS	DSVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDT
51	RVTAQLLSAY	GIQKGLYSVR	EHNERQMA DK	IVGSLSDGMV	VAQVSDAGTP
101	AVCDPGAKLA	RRVREVGFKV	VEPVVGASAVM	AALSVAAGVAG	SDFYFNQFVFP
151	PKSGERRKLF	AKWVRVAFPV	VMFETPHRIG	ATLADMAELF	PERRLMLARE
201	ITKTFETFLS	LTVGEIQTAL	AADGNQSRGE	MVLVLVPAQD	EKHEGLSESA
251	ONIMKLLTAE	LPTKQAAELA	AKITGEGKKA	LYDLALSWKN	K*

45			10	20	30	40	50	60
	orf75a.pep	M	FQKHLQKASDSVVGGTLYVVPATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY	:	:	:	:	:
	orf75-1	M	FQKHLQKASDSVVGGTLYVVPATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY	:	:	:	:	:
50			10	20	30	40	50	60
	orf75a.pep	G	IQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARREVGVGFKV	:	:	:	:	:
	orf75-1	G	IQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARREVREAGFKV	:	:	:	:	:
55			70	80	90	100	110	120
	orf75a.pep	G	IQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARREVGVGFKV	:	:	:	:	:
	orf75-1	G	IQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARREVREAGFKV	:	:	:	:	:
60			70	80	90	100	110	120
	orf75a.pep	V	PVVGASAVMAALS	VAGVAGSDFYFNGFVPPKSGERRKLF	AKWVRVAFPVVMFETPHRIG	:	:	:
	orf75-1	V	PVVGASAVMAALS	VAGVEGSDFYFNGFVPPKSGERRKLF	AKWVRAAFPIVMFETPHRIG	:	:	:
65			130	140	150	160	170	180
	orf75a.pep m	A	T	LADMAELFPERRMLLAREITKTFTETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD	:	:	:	:

-199-

orf75-1		ATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD	
		190 200 210 220 230 240	
5	orf75a.pep	250 260 270 280 290 EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX 	
	orf75-1	EKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX 250 260 270 280 290	
10 <u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>			
ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from <i>N.gonorrhoeae</i> :			
15	orf75.pep	MFVFQTAFFMFQKHLQKASDSVVGGLYVVPATPIGNLADITLRALAVLQKA----AEDTR	56
	orf75ng	MSVFQTAFFMFQKHLQKASDSVVGGLYVVPATPIGNLADITLRALAVLQKADIICAEDTR	60
20	orf75.pep	VTAQLLSAYGIQKGLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75ng	VTAQLLSAYGIQGRVSVREHNERQMADKIVIGFLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
25	orf75.pep	RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFKWWRAAFPV	176
	orf75ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKWWRAAFPVV	180
30	orf75.pep	MFETPHRIGALADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELFPERRMLLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
30	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFFM FQKHLQKASD SVVGGLYVVP ATPIGNLADI TLRALAVLQK
	51	ADIIICAEDTR VTAQLLSAYG IQGRVSVRE HNERQMADKV IGFLSDGLVV
40	101	AQVSDAGTPA VCDPGAKLAR RVREAGFKV VPVGASAVMA ALSVAGVAES
	151	DFYFNGFVPP KSGERRKLFA KWWRAAFPVV MFETPHRIGA TLADMAELFP
40	201	ERRMLLAREI TKTFTETFLSG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE
	251	KHEGLSESAQ NAMKILAAEL PTKQAAELAA KITGEGKKAL YDLALSWKNK
	301	*

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
	51	ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCAGAC ATTACCCTGC
50	101	GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATTGTGC CGAAGACACG
	151	CGCGTTACTG CGCAGCTTTT GAGCGGTAC GGCATTACAG GCAGGTGGT
55	201	CAGTGTGCGC GAACACAACG AGCGGCAGAT GGCGGACAAG GTAATCGGTT
	251	TCCTTTTCAGA CGGCCTGGTT GTGGCGCAGG TTTCCGATGC GGGTACGCCG
60	301	GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GCGAAGCAGG
	351	GTTCAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTAATG GCGGCGTTGA
65	401	GTGTGGCCGG TGTGGCGGAA TCCGATTTTT ATTTCAACGG TTTTGTACCG
	451	CCGAAATCGG GCGAACGTAG GAAATGTTT GCCAAATGGG TGCGGGCGGC
70	501	ATTTCTGTGC GTCATGTTT AAACGCGCA CCGAATCGGG GCAACGCTTG
	551	CGCATATGGC GGAATGTTC CCCGAACGCC GTCTGATGCT GGCGCGCGAA
75	601	ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTC
	651	GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTGG
80	701	TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTGTG CGAGTCTGCG
	751	CAAAATGCGA TGAAAATCCT TGCGGCGGAG CTGCCGACCA AGCAGGCGGC
85	801	GGAGCTTGCC GCCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT
	851	TGGCACTGTC GTGGAAAAAC AAATGA

60 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQGRVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALS VAGVAE SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLADMAELF PERRLMLARE
 5 201 ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

		10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAY					
	orf75ng-1	MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAY					
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf75-1.pep	GIQGKLVSVREHNERQMA DKIVGYLS DGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
	orf75ng-1	GIQGRVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKV					
		70	80	90	100	110	120
20		130	140	150	160	170	180
	orf75-1.pep	VPVVGASAVMAALS VAGVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFPV VMFETPHRIG					
	orf75ng-1	VPVVGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFPV VMFETPHRIG					
25		130	140	150	160	170	180
		190	200	210	220	230	240
	orf75-1.pep	ATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD					
30	orf75ng-1	ATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD					
		190	200	210	220	230	240
		250	260	270	280	290	
	orf75-1.pep	EKHEGLSESAQNIMKILTAELPTKQAAELA AKITGEGKKALYDLALSWKNKX					
35	orf75ng-1	EKHEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)
 40 >gi|606086 (U18997) ORF f286 [Escherichia coli]
 >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286
 Score = 218 bits (550), Expect = 3e-56
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
 45 Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITLRLALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLQLHFGIN 59
 50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVP 119
 55 Query: 124 VGASAVMAALS VAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFPV VMFETPHRIGATL 183
 G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
 Sbjct: 120 PGPCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAEPRTLIFYESTHRLDLS 179
 60 Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
 Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238
 Query: 243 HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
 E L A + +L AELP K+AA LAA+I G K ALY AL
 65 Sbjct: 239 EEDLPADALRTLALLQAEPLPKKAAALAAEIHGVKKNALYKYAL 282

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
51 TTTTGCGGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
//
651 ..... ..GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
10 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
15 201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
20 51 TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCCGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
25 301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAGACGAGCT
351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CTTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
551 AGTTTGCCGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG
30 601 GCGCAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA AAACGGTGTC
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAESQS
51 KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNR KIAEASFYAE
101 EYVRFLELSE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
40 201 GERYYLFLKS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG
251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

45 orf76.pep      10      20      30
      MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
      |||||
orf76a      10      20      30      40      50      60
      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAESQSQKPDGQAIRND
//
                                70      80      90

```

-202-

```

orf76.pep      XELVRNQLEQGLRQEKARLKDALLEENGVKPX
                |||||
orf76a      DVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLKDALLEENGVKPX
                200      210      220      230      240      250

```

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

```

1  ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51  TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
101  TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151  AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAC
201  TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251  AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301  GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
351  GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
401  GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
15  451  GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
501  TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCAGACCG CTGGCTTCGC
551  AGTTTGCAGC GATGAATCGG GCGACGTTA CCCGCGATCC GGTCAAATTG
601  GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651  CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTCGAGAC
20  701  AGGAAAAAGC CCGCTTGAAG ATCGATGCCA TTTTGAAGA AAACGGTGTC
751  AAACCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 300>:

```

1  MKQKKTAATAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
25  51  KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNR F KIAEASFYAE
101  EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151  GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
201  GERYYLFLKS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDAILEENG V
251  KP*

```

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

```

30  orf76a.pep      10      20      30      40      50      60
      MKQKKTAATAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      |||||
orf76-1      MKQKKTAATAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      10      20      30      40      50      60

35  orf76a.pep      70      80      90      100     110     120
      AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSESALRQF
      |||||
orf76-1      AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSEDELHKF
      70      80      90      100     110     120

40  orf76a.pep      130     140     150     160     170     180
      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      ||:|||||
orf76-1      YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      130     140     150     160     170     180

45  orf76a.pep      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
      |||||
orf76-1      LASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
      190     200     210     220     230     240

50  orf76a.pep      250
      IDAILEENGVKPX
      ||:|||||
orf76-1      IDALLEENGVKPX
      250

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

-203-

```

orf76.pep    MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL          30
              |||
orf76ng      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQADRHAESQSRPDGQAIRND  60
              //
5  orf76.pep                                ELVRNQLEQGLRQEKARLKIDALLEENGVPK  251
              |||
orf76ng      VTRNPVKLGERYYL FKLGA VGKNPDAQPFELVRNQLEQGLRQEKARLKIDALLEENGVPK  251

```

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

```

10      1  ATGAAACAGA AAAAGACCGC TGCCGCAGTT ATTGCTGCAA TGTGGCAGG
      51  TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
     101  TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCCGA GCAGTCCCAA
     151  AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAC
     201  TTTGGAAGTT TTGAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
     251  AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
    301  GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
    351  GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATG CAGCAGGTCA
    401  GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
    451  GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
    501  GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
    551  agttttCCGG TATGAACCGT GCGGACGTTA CCCGCAATCC GGTCAAATTG
    601  GCGCAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
    651  CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTTTGAGGC
    701  AGGAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAAGA Aacggtgtc
    751  AaacCGTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

```

1  MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAESQ
51 RPDGQAIRND AVRRLOTLEV LKNRALKEGL DKDKDVQNR F KIAEASFYAE
101 EYVRFLE RSE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR G DVTRNPVKL
201 GERYYL FKL G AVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG
251 KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

```

35      10      20      30      40      50      60
orf76-1.pep MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQADRHAESQSKPDGQAIRND
              |||
orf76ng      MKQKKTAAAVIAAMLAGFAAAKAPEIDPALVDTLVAQIMQADRHAESQSRPDGQAIRND
              |||
              10      20      30      40      50      60

40      70      80      90      100     110     120
orf76-1.pep AVRRLOTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSEDELHKF
              |||
orf76ng      AVRRLOTLEVLKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSESALRQF
              |||
              70      80      90      100     110     120

45      130     140     150     160     170     180
orf76-1.pep YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
              |||
orf76ng      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
              |||
              130     140     150     160     170     180

50      190     200     210     220     230     240
orf76-1.pep LASQFAAMNRGDVTRDPVKLGERYYL FKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
              |||
orf76ng      LASQFAGMNRGDVTRNPVKLGERYYL FKLGA VGKNPDAQPFELVRNQLEQGLRQEKARLK
              |||
              190     200     210     220     230     240

55      250
orf76-1.pep IDALLEENGVPKX
              |||
60      orf76ng IDALLEENGVPKX
              |||
              250

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
33k lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
>gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
5 >gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLESETVSE 114
VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLDDKKYKVSDEIDNKLKEYKTQLGDQYTALEKQYGYKDYLKKEQVKYELLTQ 112

Query: 115 SA-----LRQFYERQIRMIKLQOVSFATEEEARQAQQLLLKGLSFEGIMKRYPN 163
A ++++E I+ + A ++ A + ++ L KG FE L K Y
15 Sbjct: 113 KAAKDNIKVTDADIKEYWEGLKGRIRASHILVADKKTAEEVEKKLKKGEKFDLAKEYST 172

Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDA 218
D A G F Q+ E + + G+V+ DPVK Y++ K +E D
20 Sbjct: 173 DSSASKGGDLGWFAKEGQMDTFSKAAFKLKTGEVS-DPVKTQYGYHIKKTEERGKYDD 231

Query: 219 QPFELVRNQLEQGLRQEKA 237
EL LEQ L A
25 Sbjct: 232 MKKELKSEVLEQKLNDNAA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1 ATGAAAAAAT CTTTCCTTAC GCTTGTTCTG TATTCGTCTT TACTTACCGC
51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGGC
101 GCAAAAATTG CCGAAACGTT TGCGCTGACA TTTGTGATTG CTGCGCTGTA
40 151 TCTGTTTGGC CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
251 ATGACG.... //

1201 ..... CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
45 1251 ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
1351 GTTGTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTCAC
1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTCGCGAAG GCTCGGTAAC
50 1501 GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
1551 AGGCGGAATA TGT'TATCCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```

1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAAKIAETF ALTFVIAALY

```

-205-

51 LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....
 401 ...QTVFEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
 451 LYSYDPAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
 501 GNLTGDDAGS LNIRDGKAEY VYPQ*

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

1 ATGAAAAAAT CTTTCCTTAC GCTTGTCTCG TATTCGTCTT TACTTACCGC
 51 CAGCGAAAT GCCTATCGCT TTGTATTTGG GATTGAAACC TTACCGGCGG
 101 CAAAAATTGC GGAAACGTTT GCGCTGACAT TTGTGATTGC TCGCTGTAT
 151 CTGTGTTGCGC GTTATAAGGT GACGCGTTTG TTGATGCGG TGTTTTTTGC
 201 GTTCAGCATT ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
 251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
 301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCCTG TGTGTGGGG
 351 CGTGTGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
 401 CGCATTTTTC TGCCGATATA CTGTTGCCT TCCTAATGCT GATGATTTTC
 451 GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTTCG CCAAACCGAC
 501 ATACAGCCCG ATCAAAGCCA ATTATTTTCT CTTGCTTAT TTTGTCGGAC
 551 GCGTGTGGCC GTATCAGTTG TTTGATTAA GCAGGATTCC CGCCTTAAG
 601 CAGCTGCTC CAAGCAAAAT CGGCGAGGCG AGTGTCAAA ATATCGTCCT
 651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAGCTG TTTGGCTACG
 701 GACGCGAAAC TTCCCGCTTT TTAACCCGCG TGTCGCAAGC CGATTTTAAG
 751 CCGATTGTGA AACAAAGTTA TTCCGCGAGC TTTATGACTG CAGTGTCCCT
 801 GCCCAGTTTT TTCAATGCGA TACCGCACGC CAACGGCTTG GAACAAATCA
 851 GCGGCGGCGA TACCAATATG TTCCGCTCG CCAAAGAGCA GGGCTATGAA
 901 ACGTATTTTT ACAGCGCGCA GCGGAAAAC GAGATGGCGA TTTTGAACCT
 951 AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
 1001 ACGCAACCG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
 1051 AAAATCAATT TGCAGCAGG CAAGCATTTT ATCGTGTTC ACCAAGCGCG
 1101 TTCCGACGCC CCATACGCGC CATTGTTGCA GCCTCAAGAT AAAGTATTTC
 1151 GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
 1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAGCAGC CTGACGGCAA
 1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTT CGCCAAGATA
 1301 TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATCTCGT GCCGCTAGTG
 1351 TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
 1401 GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACCA
 1451 CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGC CTCGGTAACG
 1501 GGCAACCTGA TTACGGGTGA TGCAGGCAGC TTGAACATTC GCGACGGCAA
 1551 GCGGAATAT GTTTATCCGC AATGA

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

40 1 MKKSFLTLVL YSSLLTASEI AYRFVFGIET LPAAKIAETF ALTFVIAALY
 51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYW LMLKEVTEVG
 101 SAGASMLDKL WLPVLWGVLE VMLFCSLAKF RKRTHFSADI LFAFLMLMIF
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLRSIPAFK
 45 201 QPAPSKIGQG SVQNIIVLIMG ESESAHLKL FGYGRETSPP LTRLSQADFK
 251 PIVKQSYSAG FMTAVSLPSF FNAIPHANGL EQISGGDTNM FRLAKEQGYE
 301 TYFYSAQAEN EMAILNLIGK KWIDHLIPT QLGYGNGDNM PDEKLLPLFD
 351 KINLQQKHF IVLHQRGSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
 401 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
 451 LYSYDPAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
 50 501 GNLTGDDAGS LNIRDGKAEY VYPQ*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

55 orf81.pep 10 20 30 40 50 60
 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAAKIAETFALT FVIAALYLFARNKVTRL
 orf81a 10 20 30 40 50 60
 MKKSFLVLF YSSLLTASEI AYRFVFGIET LPAAKMAETFALT FVIAALYLFARYKATRL
 60 70 80

```

orf81.pep      LIAVFFAFSIIANNVHYADYQSWMT
                |||||
orf81a         LIAVFFAFSIIANNVHYAVYQSWITGINYWLMKEITEVGGAGASMLDKLWLPALWGVLE
                70      80      90      100      110      120
                //
                                120      130      140
orf81.pep      QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
orf81a         IPHANGLEQISGGDIVDKYDNTIHKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD
                280      290      300      310      320      330
                150      160      170      180      190      200
orf81.pep      IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
                |||||
orf81a         IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG
                340      350      360      370      380      390
                210      220      230
orf81.pep      CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
                |||||
orf81a         CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
                400      410      420

```

1	ATGAAAAAAT	CCCTTTTCGT	TCTCTTTCTG	TATTCGTCCC	TACTTACTGC
51	CAGCGAAAT	GCTTATCGCT	TTGTATTCCG	AATTGAAACC	TTACCGCGCT
101	CAAAAAATGGC	AGAAACGTTT	CGCGTGACAT	TTGTGATTGC	TGCGCTGTAT
151	CTGTTTGC	GTTATAAGGC	AACGCGTTTG	TTGATTCGCG	TGTTTTTCGC
201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGA
251	TAACGGGCAAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
301	GCGCGAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
351	CGGTGTTGAA	GCTGATGTTG	TTTGACGCTT	TGCCAAGTTC	CGCCGTAAGA
401	CGCATTTTTT	TGCCGATATA	CTGTTTGCTT	TCCTAATGCT	GATGATTTTT
451	GTGCGTTTCG	TCGACACGAA	ACAAGAACAC	GGTATTTCGC	CCAAACCGAC
501	ATACAGCCGC	ATCAAGCCCA	ATTATTTCAG	CTTCGGTTAT	TTTGTCCGAC
551	GCGTGTGGCC	GTATCAGTTG	TTTGATTTAA	GCAAGATTCC	TGTGTTCAAA
601	CAGCCTGCTC	CAAGCAGAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCCG
651	GATTATGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAAATTG	TTTGGCTACG
701	GGCGCGAAAC	TTCCGCGTTT	TTGACCCAGC	TTTCCGAAGC	GAGTTTTAAG
751	CCGATTGTGA	AACAAAGTTA	TTCCGCGAGC	TTTATGACGG	CAGTATCCCT
801	GCCAGTTTTT	TTTAACTGCA	TACCGCATGC	CAACGGCTTG	GAAACAAATCA
851	GCGGCGGCGA	TATTCGTGAT	AAGTACGACA	ACACCATCCA	CAACACCGAC
901	CAAAATGATT	AAACCTGATT	CAGAGCAGTG	CAAAAGCAGC	CTGACGGCAA
951	CTGGCTGTTT	GCCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
1001	TCTACAATCA	AGGCCAGGGT	CAGCCCGACA	GCTATCTCGT	GCCGCTGGTG
1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	CGTGCCAACC	AGGCTTTTGC
1101	GCCTTGGCAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTCACT
1151	CGTTGGGCTA	CGATATGCCG	GTTTCCAGGT	GTCGCGAAGG	CTCGGTAAAC
1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATTG	GCGACGGCAA
1251	GGCGGAATAT	GTTTATCCGC	AATGA		

55

1	MKKSFLVFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFFVIAALY
51	<u>LFARYKATRL</u>	<u>LIAVFFAFSI</u>	IANNVHYAVY	QSWITGINYW	LMLEKETEVEG
101	GAGASMLDKL	WLPALWGVLE	VMLFCSLAKF	RRKTHFSADI	LFAFLMLMFI
151	VRSFDTQKEH	GISPKPTYSR	IKANYFSFGY	FVGRVLFPYQL	FDLSKLMPVFK
201	QPAPSRIGQG	SIQNIVLIMG	ESESAHLKL	FGYGRETSPF	LTQLSQADFK
251	PIVKQSYSAG	FMTAVSLPSF	FNVIPHANGL	EQISGGDIVD	KYDNTIHKTD
301	QMIQTVEEQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNQGTV	QPDSYLVPLV
351	LYSPDKAVQQ	AAHQAFAPCE	IAFHQQQLSTF	LIHTLGYDMP	VSGCREGSVT
401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

```

              10      20      30      40      50      60
orf81a.pep    MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALT FVIAALYLFARYKATRL
               |||:::|| | ||||| ||||| ||||| ::||| ||||| ||||| :||
65 orf81-1     MKKSFLT LVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALT FVIAALYLFARYKVTRL
               10      20      30      40      50      60

```

-207-

5	orf81a.pep	70 80 90 100 110 120	LIAVFFAFSIIANNVHYAVYQSWITGINYWMLLKEITEVGGAGASMLDKLWLPALWGVLE
	orf81-1	70 80 90 100 110 120	LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLLKEVTEVGSAGASMLDKLWLPVWLVLE
10	orf81a.pep	130 140 150 160 170 180	VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
	orf81-1	130 140 150 160 170 180	VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY
15	orf81a.pep	190 200 210 220 230 240	FVGRVLPYQLFDLSKIPVKQPAPSRIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPF
	orf81-1	190 200 210 220 230 240	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRGRETSPF
20	orf81a.pep	250 260 270 280	LTQLSQADFKPIVKQSYSGFMTAVSLPSFFNVI PHANGLEQISGGD-----
	orf81-1	250 260 270 280 290 300	LTRLNQADFKPIVKQSYSGFMTAVSLPSFFNAI PHANGLEQISGGDTNMFRLAKEQGYE
30	orf81a.pep		-----
	orf81-1	310 320 330 340 350 360	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF
35	orf81a.pep	290 300 310 320	-----IVDKYDNTIHKTQMIQTVFEQLQKQPDGNWLF
	orf81-1	370 380 390 400 410 420	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTQMIQTVFEQLQKQPDGNWLF
40	orf81a.pep	330 340 350 360 370 380	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
	orf81-1	430 440 450 460 470 480	AYTSDHGQYVRQDIYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF
45	orf81a.pep	390 400 410 420	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
	orf81-1	490 500 510 520	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX

50 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

55	orf81.pep	MKKSFLTIVLYSSLLTASEIAYPLELGIETLPAAKIAETFALT FVIAALYLFARNKVTRL	60
	orf81ng	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL	60
60	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
	orf81ng	LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLLKEVTEVGSAGASMLDKLWLPALWGVAE	120
65	orf81.pep	QTVFEQLQKT PDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPQDKVFGEADIVDKYDNTIHKTQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD	433
65	orf81.pep	IYNQGTVPDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81ng	IYNQGTVPDSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493


```

orf81.pep    CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ  524
             |||||:|||||
orf81ng      CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ  524

```

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

```

5      1  ATGAAAAAAT CCCTTTTCGT TCTCTTCTG TATTCATCCC TACTTACCGC
      51  CAGCGAAATC GCCTATCGCT TTGTATTCGG AATTGAAACC TTACCGGCTG
     101  CAAAAATGGC GGAAACGTTT GCGCTGACAT TTATGATTGC TGCGCTGTAT
     151  CTGTTTGCGC GTTATAAGGC TTCGCGGCTG CTGATTGCGG TGTTTTTCGC
     201  GTTCAGCATG ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
    10  251  TGACGGGTAT TAACTATTGG CTGATGCTGA AAGAGTTAC CGAAGTCGGC
     301  AGCGCGGGCG CGTCGATGTT GGATAAGTTG TGGCTGCCTG CTTTGTGGGG
     351  CGTGGCGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTT TGCCGATATA CTGTTTGCTT TCCTAATGCT GATGATTTTC
     451  TGCGGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTCGC CCAAACCGAC
    15  501  ATACAGCCGC ATCAAAGCCA ATTATTTCAG CTTCGGTTAT TTGTGCGGGC
     551  GCGTGTGGCC GTATCAGTTG TTTGATTAA GCAAGATCCC TGTGTTCAAA
     601  CAGCTGTGCT CAAGCAAAAT CGGGCAAGGC AGTATTCAAA ATATCGTCCT
     651  GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGGTTACG
     701  GCGCGGAAAC TTCGCGGTTT TTAACCGGC TGTCGCAAGC CGATTTTAAG
    20  751  CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACGG CAGTATCCCT
     801  GCCCAGTTTC TTTAAGTCA TACCGCAGGC CAACGGCTTG GAACAAATCA
     851  GCGGCGGCGA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
     901  ACGTATTTTT ACAGTGCCCA GGCTGAAAAC CAAATGGCAA TTTTGAACCT
     951  AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
    25 1001  ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
    1051  AAAATCAATT TGCAGCAGGG CAGGCATTTT ATCGTGTTCG ACCAACGCGG
    1101  TTCGACGCCC CCATACGGCG CATTGTTGCA GCCTCAAGAT AAAGTATTCG
    1151  GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
    1201  CAAATGATTC AAACCTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
    30 1251  CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTG CGCCAAGATA
    1301  TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATATTGT GCCTCTGGTT
    1351  TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
    1401  GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACAC
    1451  CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACA
    35 1501  GGCAACCTGA TTACGGGCGA TGCAGGCAGC TTGAACATTC GCAACGGCAA
    1551  GCGGAATAT GTTTATCCGC AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 310>:

```

      1  MKKSLFVLFL YSSLLTASEI AYRFVFGIET LPAAKMAETF ALTFMIAALY
     51  LFARYKASRL LIAVFFAFSM IANNVHYAVY QSWMTGINYW LMLKEVTEVG
    101  SAGASMLDKL WLPALWGVAE VMLFCSLAKF RRKTHFSADI LEAFMLLMIF
    151  VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLISKIPVFK
    201  QPAPSKIGQG SIQNIVLIMG ESESAHLKL FGYGRETSPP LTRLSQADFK
    251  PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDTNM FRLAKEQGYE
    301  TYFYSAQAEQ QMAILNLIGK KWIDHLIQPT QLGYNNGDNM PDEKLLPLFD
    351  KINLQQRHFH IVLHQRGSHA PYGALLQPDQ KVFGEADIVD KYDNTIHKTD
    401  QMIQTVEEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTQ QPDSYIVPLV
    451  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
    501  GNLITGDAGS LNIRNGKAEY VYPQ*

```

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

```

50      10      20      30      40      50      60
orf81ng-1.pep MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL
             |||||:|||||
orf81-1        MKKSFLTLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALTFVIAALYLFARYKVTRL
             |||||:|||||
55      70      80      90     100     110     120
orf81ng-1.pep LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLLKEVTEVGSAGASMLDKLWLPALWGVAE
             |||||:|||||
orf81-1        LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLLKEVTEVGSAGASMLDKLWLPVWLVLE
             |||||:|||||
60      70      80      90     100     110     120
orf81ng-1.pep VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRKANYFSFGY
             |||||:|||||
orf81-1        VMLFCSLAKFRRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRKANYFSFGY
             |||||:|||||
65

```

-209-

		130	140	150	160	170	180
5	orf81ng-1.pep	190	200	210	220	230	240
	orf81-1	190	200	210	220	230	240
10	orf81ng-1.pep	250	260	270	280	290	300
	orf81-1	250	260	270	280	290	300
15	orf81ng-1.pep	310	320	330	340	350	360
	orf81-1	310	320	330	340	350	360
20	orf81ng-1.pep	370	380	390	400	410	420
	orf81-1	370	380	390	400	410	420
25	orf81ng-1.pep	430	440	450	460	470	480
	orf81-1	430	440	450	460	470	480
30	orf81ng-1.pep	490	500	510	520		
	orf81-1	490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi 1256380 (U50906) outer membrane adherence protein-associated protein [<i>E. coli</i>] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMVQRLLFFVLTLIVVKRISSLPRLLLVAAPFVL-LTAADMSISLY- 86
50	Query: 82 SWMT-----GINYWMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLFCSLAKFRKRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLAVI IKYDV 141
55	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
60	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPHYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
65	Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSLP---SFFNVIPHANGLEQISGGDTNMFLAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPLSLTADSVLSH-----DIHNYPDNIINMANQAG 310
70	Query: 299 YETFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLLPPLSQUALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPQDKVFEADIVDK-YDNTIHKTDQMIQTVEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPQDDQDACYDNSIHYTDSLGLQVFEKLLK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPDSYIVPL-VLYSP 454
 D Y +DHG ++++Y G +Y VP+ + YSP
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVPFMIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1  ...ACCCTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGGC GCACACTGAC
51  CGGCATACTC GCCCcCGGCG GCGGCAAAACG CTTTGCCGTC GAACAAGAAC
101  TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTTGTCCGCC
151  yTAAAGGAC GCAAAGCCGC CyTTTACGTC TCCGTTATGG GCGACCAAGG
201  TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251  GCGGCTACCA CAACAACCCC GAAAGTGCCA CCAATACAG CTACCCCGCC
301  TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCGTAACCAC
351  TTCCACATCG CTTTGAACG CCCCCGCCGC CGyCyTGACG AAAAAACGCG
401  GACGCAAAGG CGAACGcTCC GCCGGACTGT CCGTCAACGG CACGGGCGAC
451  TACCGCAACG AAACCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
501  CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
551  CCGrATACGC CGACACCGAC GTATTCGTAA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

1  ..TLLFIPLVL TXCGTLTGIL AHGGGKRFV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQYSYPA
101  YDTTATTKSD ALSSVTSTTS LLNAPAAXLT KNSGRKGRS AGLSVNGTGD
151  YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

1  ATGAAACCC TGTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51  ACTGACCGGC ATACCCGCC ACGGCGGCGG CAAACGCTT GCCGTCGAAC
101  AAGAAGCTCG CGCGCATCG TCCCGCGCCG CCGTCAAAGA AATGGATTG
151  TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
201  CCAAGGTTTC GGCAACATAA GCGGCGGACG CTAATCTATC GACGCACTGA
251  TACGCGGCGG CTACCACAAC AACCCTGAAA GTGCCACCCA ATACAGCTAC
301  CCGCCTACG ACATACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
351  AACCACTTCC ACATCGCTT TGAACGCCCC CGCCGCGGCC CTGACGAAAA
401  ACAGCGGACG CAAAGGCGAA CGCTCCGCC GACTGTCCGT CAACGGCAGC
451  GCGGACTACC GCAACGAAAC CCTGCTCGCC AACCCCGCG ACGTTCCTT
501  CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551  TACCGCCCGA ATACGCGGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
601  GGCACCGTCC GCAGCCGTAC CGAAGTGCAC CTCTACAACG CCGAAACCCT
651  TAAAGCCCAA ACCAAGCTCG AATATTTTCG CGTTGACCGC GACAGCCGGA
701  AACTGCTGAT TACCCTAAA ACCGCGGCT ACGAATCCCA ATACCAAGAA
751  CAATACGCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
801  CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
851  CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC
901  GATGTCGGCA ACGAAGTCAT CCGCGCGCG AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

1  MKTLLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101  PAYDTTATTK SDALSSVTTS TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
151  GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVEPEYAD TDVFVTVDVF
201  GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
251  QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNKKPK
```

301 DVGNEVIRRR KGG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N.*

5 *meningitidis:*

		10	20	30	40	50	
	orf83.pep	<u>TLLLFIPVLVLTXC</u> GTLTGILAHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAX :					
10	orf83a	<u>MKTLLXLIPLVLTAC</u> GTLTGP AHGGGKRFAVEQELVAASSRAAVKEMDLSALKGRKAAL 10 20 30 40 50 60					
		60	70	80	90	100	110
	orf83.pep	YVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSALSSVTTS 					
15	orf83a	YVSVMGDQGSGNISGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSALSSVTTS 70 80 90 100 110 120					
		120	130	140	150	160	170
20	orf83.pep	TSLLNAPAAXLTKNSGRKGERSAGLSVNNGDYRNETLLANPRDVSFLTNLIQTVFYLRG 					
	orf83a	TSLLNAPAAALTKN SGRKGERSAGLSVNNGDYRNETLLANPRDVSFLTNLIQTVFYLRG 130 140 150 160 170 180					
		180	190				
25	orf83.pep	IEVVPPXYADTDVFVTVDV 					
	orf83a	IEVVPEYADTDVFVFGTVRSRTLHLHYNAETLK AQT KLEYFAVD RDSRKLLIAPK 190 200 210 220 230 240					

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAACCC	TGCTCTCTCT	CATCCCCCTC	GTCCTCACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCCC	ACGGCGCGCG	CAAACGCTTT	GCCGTGCAAC
	101	AAGAATCTCGT	CGCGCATCG	TCCCGCGCGG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGAA	AGCCGCCCTT	TACGTCTCGG	TTATTGGCGCA
35	201	CCAAGGTTTCG	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACTGA
	251	TACGCGCGCG	CTACCACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAG
	301	CCCGCTTACG	ACACTACCGC	CACCACCAA	TCCGACGCGC	CTCTCCAGCT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCGGCC	TGTGACAAAA
40	401	ACAGCGGACG	CAAAGGCGAA	CGCTCCGCGG	GACTGTCCGT	CAACGGCACG
	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCGG	ACGTTTCTCT
	501	CCTGACCAAC	CTCATCCAA	CCGTCTTCTA	CCTGCGCGCG	ATCGAAGTCG
	551	TACCGCCCGA	ATACCCGCGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
45	601	GGCACCCTCC	GCAGCCGAC	CGAAGTCAC	CTCTACAACG	CCGAAACCTT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTCGC	CGTTGACCGC	GACAGCCCGGA
	701	AACTGCTGAT	TGCCCTTAA	ACCGCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCC	TCTGGATGGG	ACCTTACAGC	GTCGGCAAAA	CCGTCAAAGC
	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAAACCC
	901	GATGTCGGCA	ACGAATCAT	CCGCGCGCGC	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLLXLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGGRYSI	DALIRGGYHN	NPEASATQSY
	101	PAYDTTATTK	SDALSSVTS	TSLLNAPAAA	LTKNSGRKGE	RSAGLSVNGT
	151	GDYRNETLLA	NERDVSFLT	LIQTVFYLRG	IEVVPPEYAD	TDVFVTVDFV
	201	GTVRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLLTAPK	TAAYESQYQE
55	251	QYALWMGPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNRP	DFKQNNGGKP
	301	DVGNEVIRRR	KGG*			

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

orf83a.pep 10 20 30 40 50 60
MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL

10

15

20

25

30

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

40

45

50

The complete length ORF83ng nucleotide sequence <SEQ ID 317> is:

55

60

65

-213-

5
 651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTCGACCGC GACAGCCGGA
 701 AACTGCTGAT TGCCCTAA ACGCCGCCT ACGAATCCCA ATACCAAGAA
 751 CAATACGCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC
 801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCC CACGGCGACA
 851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCC
 901 GATGTCGGCA ACGAAGTCAT CCGCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10
 1 MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
 51 SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPDSATRYSY
 101 PAYDTTATTK SDALSGVTTTS TSLNAPAAA LTKNNGRKGE RSAGLSVNGT
 151 GDYRNETLLA NPRDVSFLTNI LIQTVFYLRG IEVVPPEYAD TDVFVTVDFV
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
 251 QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRPFDFKQNNKKNP
 301 DVGNEVIRRR KGG*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
orf83-1.pep		MKTLLLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL	SALKGRKAAL		
20 orf83ng		MKTLLLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL	SALKGRKAAL		
		10	20	30	40	50	60
		70	80	90	100	110	120
orf83-1.pep		YVSMGDQGS	GNISGGRYSIDALIRGGYHNPESATQYSYPAYDTTATTKSDALSSVTTS				
25 orf83ng		YVSMGDQGS	GNISGGRYSIDALIRGGYHNPDSATRYSYPAYDTTATTKSDALSGVTTS				
		70	80	90	100	110	120
		130	140	150	160	170	180
orf83-1.pep		TSLNAPAAA	LTKNSGRKERSAGLSVNGTG	DYRNETLLANPRDVSFLTNI	LIQTVFYLRG		
30 orf83ng		TSLNAPAAA	LTKNNGRKERSAGLSVNGTG	DYRNETLLANPRDVSFLTNI	LIQTVFYLRG		
		130	140	150	160	170	180
		190	200	210	220	230	240
orf83-1.pep		IEVVPPEYADTDVFVTVDFG	TVRSRTELHLYNAETLKAQ	TKLEYFAVDR	DSRKLLITPK		
40 orf83ng		IEVVPPEYADTDVFVTVDFG	TVRSRTELHLYNAETLKAQ	TKLEYFAVDR	DSRKLLIAPK		
		190	200	210	220	230	240
		250	260	270	280	290	300
orf83-1.pep		TAAYESQYQE	QYALWGTGPKVSKTVKASDR	LMVDFSDITPYGDTTAQNRPFDFKQNNKKNP			
45 orf83ng		TAAYESQYQE	QYALWMGPYSVGKTVKASDR	LMVDFSDITPYGDTTAQNRPFDFKQNNKKNP			
		250	260	270	280	290	300
		310					
orf83-1.pep		DVGNEVIRRRKGGX					
50 orf83ng		DVGNEVIRRRKGGX					
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
5  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
10 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
15 551 AAGTTTATGA CTGTATsrr TmmGCGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
651 CGTGTTTTGT GGCCTGTCCT ATAAAATGTT GagCaGTTAC GGAAAAAAC
701 aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
20 801 AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGrAAGC AAGCcgaTTT
851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
951 gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AAacGGCTTG
25 1001 CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGCAGGAAG TTCAGCAAAG
1051 CGCGCAgCAA CATTGGGACA GGGCGcCAAG TTGCCACATT GGGCGGAAAA
1101 CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAACGCG GGAAACCGTT
1151 TGAAGGAATC GGaCGGGGGC GTGGTCCGAT CGGCAAAC TG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

1  MAEICLITGT PGSGKTLKLV SMMANDEMFK PDEKAIRRV FTVNIKGLKIP
30 51  HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRHQG IDIFVLQGP KLLDQNLRTL VRKHYHIASN
151 KMGMRITLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNVK
201 KRSKWFTLTP VIVLLIPVFV GLSYKMLSSY GKQKEEPAAQ ESAATEQQAV
251 LPDKTEGEPV NNGNLTADMV VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
35 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGQEVQQS
351 AQQHSDRAQV ATLGGKPXQN LMYDNWEERG KPFEGIGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
40 51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
45 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTTGAC TCAAGGTCCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
55 551 AAGTTTATGA CTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCa GTAATAGTAT TGCTGATTCC
651 CGTGTTTTGT GGCCTGTCCT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
55 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
951 GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
1001 CGTTTAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
60 1051 CGCGACGAAC ATTCGGACAG GCGCAAGTT GCCACATTGG CGGAAAAACC
1101 GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

```

1  MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP
51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK
201 KRSKWFYTLV VIVLLIPVFV GLSYKMLSSY GKKQEEPAQA ESAATEQQAV
251 LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQQS
351 AQQHSRAQV ATLGKGP*QN LMYDNWEERG KPFEGIGGGV VGSAN*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N.*

meningitidis:

15	orf84.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK
	orf84a	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRKVFTNIKGLKIPHTYIETDAKK
20	orf84.pep	LPKSTDEQLSAHDMEYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG
	orf84a	LPKSTDEQLSAHDMEYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG
25	orf84.pep	IDIFVLTQGP KLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIYT
	orf84a	IDIFVLTQGS KLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIYT
30	orf84.pep	LDKKVYDLYXAEVHTVNVKRSKWFYTLVPIVLLIPVFVGLSYKMLSSYGKKQEEPAQA
	orf84a	LDKKVYDLYESA EVHTVNVKRSKWFYTLVPIVILLIPVFVGLSYKMLSSYGKKQEEPAQA
35	orf84.pep	ESAATEQQAVLPDKTEGEPVNNNGNLTADMVPTLSEKPKSKPIYNGVRQVRTFEYIAGCI
	orf84a	ESAATEHQAVFQDKTEGEPVNNNGNLTADMVPTLSEKPKSKPIYNGVRQVRTFEYIAGCV
40	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGEVQQSAQQHSRAQV
	orf84a	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDQQSEQHHSRDPQV
45	orf84.pep	ATLGKGPXQNLMYDNWEERGKPFEGIGGGVVGVSANX
	orf84a	ATLGKGPWQNLMYDNWQERGKPFEGIGGGVVGVSANX

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

```

55 1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAAACATT
51  AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAAA
101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG
151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
60 251 TCGGGTCTAT TGTCAATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTGGAC TCAAGGCTCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGTTCAAAC
451 AAGATGGGTA TGCGTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC

```


-216-

5 501 CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
 601 AAGCGGTCAA AATGGTTTTA TACTCTGCCA GTAATAATAT TGCTGATTCC
 651 CGTTTTTGTG GGCCTGTCCT ATAAAATGTT AAGTAGTTAT GGAAAAAAC
 701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA TCAGGCAGTA
 751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACACGGTA ACCTTACCGC
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
 901 GAAGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGCATT
 10 951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
 1001 CGTTTAACCC ATATAAGAA GAAAGCCAAG GGCGGGATGT CCAGCAAAGT
 1051 GAGCAGCACC ATTCGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC
 1101 GTGGCAAAAT CTTATGTATG ATAATTGGCA GGAGCGCGGA AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRVK FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIIY LDKKVYDLYE SAEVHTVNVK
 20 201 KRKSWFYTLV VIILLIPVVFV GLSYKMLSSY GKKQEEPAQ ESAATEHQAV
 251 FQDKTEGEPV NNGNLTADMV VPTLSEKPEK KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS
 351 EQHSDRPQV ATLGGKPWQN LMYDNWQERG KPFEGIGGGV VGSAN*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25		10	20	30	40	50	60
	orf84a.pep	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRVFTNIKGLKIPHTYIETDAKK					
	orf84-1	MAEICLITGTPGSGKTLKMVSMMANDEMFKPDENGIRRVFTNIKGLKIPHTYIETDAKK					
		10	20	30	40	50	60
30		70	80	90	100	110	120
	orf84a.pep	LPKSTDEQLSAHDMEYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
	orf84-1	LPKSTDEQLSAHDMEYEWIKKPENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG					
35		70	80	90	100	110	120
		130	140	150	160	170	180
	orf84a.pep	IDIFVLTQGSKLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIIY					
	orf84-1	IDIFVLTQGPKLLDQNLRTLVRKHYHIASNKMGMRITLLEWKICADDPVKMASSAFSSIIY					
40		130	140	150	160	170	180
		190	200	210	220	230	240
	orf84a.pep	LDKKVYDLYESADEVHTVNVKVRKSWFYTLVPIILLIPVVFVGLSYKMLSSYGKKQEEPAQ					
	orf84-1	LDKKVYDLYESADEVHTVNVKVRKSWFYTLVPIVLLIPVVFVGLSYKMLSSYGKKQEEPAQ					
45		190	200	210	220	230	240
		250	260	270	280	290	300
	orf84a.pep	ESAATEHQAVFQDKTEGEPVNNNGNLTADMVPTLSEKPEKPIYNGVRQVRTFEYIAGCV					
	orf84-1	ESAATEQQAVALPDKTEGEPVNNNGNLTADMVPTLSEKPEKPIYNGVRQVRTFEYIAGCI					
50		250	260	270	280	290	300
		310	320	330	340	350	360
	orf84a.pep	EGGRTGCTCYSHQGTALKEITKEMCKDYARNGLPFNPYKEESQGRDVQQSEQHSDRPQV					
	orf84-1	EGGRTGCACYSHQGTALKEITKEMCKDYVKNGLPFNPYKEESQGEVQQSAQQHSDRAQV					
55		310	320	330	340	350	360
		370	380	390			
	orf84a.pep	ATLGGKPWQNLMYDNWQERGKPFEGIGGGVGSANX					
	orf84-1	ATLGGKPQNLMYDNWEERGKPFEGIGGGVGSANX					
60		370	380	390			
65		370	380	390			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N. gonorrhoeae*:

5	orf84.pep	MAEICLITGTPGSGKTLKVMSSMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKVMSSMANDEMFKPDENGVRKVFTNIKGLKIPHTHETDAKK	60
	orf84.pep	LPKSTDEQLSAHDMYEWIKKPENIGSIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHQG	120
10	orf84ng	LPKSTDEQLSAHDMYEWIKK PENVGAIVIVDEAQDVWPARSAGSKI PENVQWLNTHRHQG	120
	orf84.pep	IDIFVLTQGP KLLDQNLRLTLVRKHYHIA SNKMGRTLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGP KLLDQNLRLTLVRKHYHIA ANKMGRLTLLEWKVCADDPVKMASSAFSSIYT	180
15	orf84.pep	LDKKVYDLYXAEVHTVNKVKRSKWFTLPVIVLLIPVFGLSYKMLSSYGKKQEEPAAQ	240
	orf84ng	LDKKVYDLYSAE IHTVNKVKRSKWFTLPVILLIPVFGLSYKMLGSYGKKQEEPAAQ	240
20	orf84.pep	ESAATEQQAVLPDKTEGEPVNNGNLTADMVPTLSEKPKSPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAATEQQAVLPDKTEGESVNNGNLTADMVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
25	orf84.pep	EGGRTGCACYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLPFNPYKEESQGQEVQSSAQQHSRAQV	360
30	orf84.pep	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVGSAN	395
	orf84ng	ATLGGKPKQNLMYDNWEERGKPFEGIGGGVGSAN	395

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

	1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTACG	GGAAAACATT
	51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAAG	CCAGATGAAA
35	101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
	151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
	201	ACAGCTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
	251	tcggcgCAAT	CGTTATGTGC	GATGAGGCGC	AAGACGTATG	GCCCGCACGC
	301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
	351	GCATCAGGGC	ATAGATATAT	TTGTATTGAC	ACAAGGTCCT	AAACTCTTAG
40	401	ATCAGAACTT	GCGAACATTTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
	451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
	501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
	551	AAGTTTATGA	CTGTACGAA	TCCGCAGAAA	TTACACCGGT	AAACAAAGTC
	601	AAGCGTTCAA	AATGGTTTGA	TGCATTGCCC	GTATCATAT	TATTGATTCC
45	651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAC
	701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	GCAGGCAGTA
	751	CTTCCGGATA	AAACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
	801	AGATATGTTT	GTTCGACAT	TGCCCCGAAA	ACCCGAAAGC	AAGCCGATTT
	851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
50	901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
	951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
	1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
	1051	GCGCAGCAAC	ATTCGGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAACC
	1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG
55	1151	AAGGAATCGG	CGGGGCGCTG	GTCGGATCGG	CAAACCTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

	1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGVRKV	FTNIKGLKIP
	51	HTHETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIVIV	DEAQDVWPAR
60	101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLLDQNLRLTL	VKRHYHIAAN
	151	KMGLRTLEW	KVCADDPVKM	ASSAFSSIYT	LDKKVYDLYE	SAE IHTVNKV
	201	KRSKWYALP	VIILLIPLFV	GLSYKMLGSY	GKKQEEPAAQ	ESAATEQQAV
	251	LPDKTEGESV	NNGNLTADMV	VPTLPEKPES	KPIYNGVRQV	RTFEYIAGCI
	301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPYKE	ESQGQEVQSS
	351	AQQHSDRAQV	ATLGGKPKQN	LMYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEICLITGTPGSGKTLK	MVSMMA	NDEMFKP	DENGIRRKVFT	NIKGLKIPHTYIET	DAKK
5	orf84ng	MAEICLITGTPGSGKTLK	MVSMMA	NDEMFKP	DENGVR	RRKVFTNIKGLKIPHTIET	DAKK
		10	20	30	40	50	60
	orf84-1.pep	LPKSTDEQLSAHDMYEWIKK	PENIGSIV	IVDEAQDV	WPARSAGS	KIPENVQWLN	THRHQG
10	orf84ng	LPKSTDEQLSAHDMYEWIKK	PENVGAI	IVDEAQDV	WPARSAGS	KIPENVQWLN	THRHQG
		70	80	90	100	110	120
	orf84-1.pep	IDIFVLTQGP	KLLDQNL	RTRLV	RKH	YHIA	SNKMG
15	orf84ng	IDIFVLTQGP	KLLDQNL	RTRLV	RKH	YHIA	ANKMGL
		130	140	150	160	170	180
	orf84-1.pep	LDKKVYDLYESA	EVHTVN	KVKRS	KWFY	TL	PVIVLLIP
20	orf84ng	LDKKVYDLYESA	EIH	TVN	KVKRS	KWFY	ALPVI
		190	200	210	220	230	240
	orf84-1.pep	ESAAEQQAVLP	DKTEGE	PVNN	GNLT	ADM	FVPTLSEK
25	orf84ng	ESAAEQQAVLP	DKTEGE	SVNN	GNLT	ADM	FVPTLPEK
		250	260	270	280	290	300
	orf84-1.pep	EGGRTGCAC	YSHQGT	ALKEV	TELMCK	DYV	KNGLP
30	orf84ng	EGGRTGCTC	YSHQGT	ALKEV	TELMCK	DYV	KNGLP
		310	320	330	340	350	360
	orf84-1.pep	ATLGGK	XPQN	LMYD	NWE	ERGK	PFEGIG
35	orf84ng	ATLGGK	XPQN	LMYD	NWE	ERGK	PFEGIG
		370	380	390			
	orf84-1.pep	ATLGGK	XPQN	LMYD	NWE	ERGK	PFEGIG
40	orf84ng	ATLGGK	XPQN	LMYD	NWE	ERGK	PFEGIG

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

50	1	GTGGTTTTC	TGAATGCCGA	CAACGGGATA	TTGGTTCAGG	ACTTGCCTTT
	51	TGAAGTCAAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAT	ACGGGTATGC
	101	CGCGTGATTT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTTAT	CAGGCGAGTT	TTGCCGACGG	CGGTTCCGGAT	TTGACATTCA
	251	AGGCGTGGAA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTCGT	GTTGAAGGCA
55	301	ACATCCATAC	ACCACTTTCC	GTTGGAAATT	GGCAAACACA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACCT	CTATGAATGT	GGAGGACATG	AGCGAGGGCG
	401	CGGAACGGGA	AAAAAGCCTG	AAATCCACGC	TGCCCGATGT	CCGCGCGGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...TACCG
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTGCA	ATATAAAAAC	TATATGCTGC
60	551	CGGTTTTCGA	GGAACAGGAT	TATTTTTCGA	TTACCGGCAC	GCGCAGCGC.

5
10
1101

```

601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCGTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAAAACACG CTGAACATCT TTGCACAAAA
801 AGGCTATTTG GGATTGGACG AATTTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTGGG TTTGCAGATG ACCCGTTCCC C.GGTCCGCT
1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

15
20

```

1 MVFLNADNGI LVQDLPEFEVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITY QASFADGGSD LTFKAWNLGD ASREPVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRAV
151 TOEGHKYTNX XXXXXYRIRD APGQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSN I PKEQQDKMQG YFYEMLYGVM
301 NAALDETXTR YGLPEWQODE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
351 SEVRSSGLQM TRSXGPLLVY L...
```

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

25
30
35
40
45
50
55
60

```

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCCTC CGTGGTTCGC
51 TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATCGCCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTTGCCA
401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGTCCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCCTG GCGGGGTGA
551 TAGACAGTAA CCTGCTGTTG AAAGTGGGTA TGCTGACCGG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
35 651 GGGTGCGTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCCTC TGTGGTTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCTTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
40 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCCGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC
1001 TGTTGAAGGC AACATCCATA CACCAAGTTT CGTTGGAAT TGGAACAAC
1051 AAATATCGTC TTGAGTTCTGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
45 1151 TCCGCGCCGT TACTCAGGAA GGTAATAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TCGGCGAGGG CAGGCGGTGC AATATAAAAA
1251 CTATATGCTG CCGGTTTTCG AGGAACAGGA TTATTTTGG ATTACCGGCA
1301 CGCGCAGCGG CTTCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
50 1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GCGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
55 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTCAGAT GACCGGTTCC
1801 CCGGGTGC GC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
60 1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGAGCGA ACGGGATTG
1951 CAGAAGGAAT TTCCAACAAC CGTCGAGAGT CTGCAACGGC TCGCAAGGA
2001 CTTGAATCAT GACTGA
```

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

65

```

1 MSKSRSPPL LSRPWAFFS SMRFAVALLS LLGIASVIGT VLQONQPTD
51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
```

-220-

101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQKKTINRE
 151 DGSVLIAAKK GTMKNWGYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
 201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
 5 251 DLPFEVKKKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
 301 LHGITYIQAS FADGSDLTG KAWNLDASR EPVVLKATSI HQFPLEIGKH
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
 401 IVYRIRDAAG QAVEYKNYML PVLQEQQDYFW ITGTRSGLQQ QYRWLRIPLD
 451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL
 10 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
 601 PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFS DGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

	orf88.pep				10	20	30
					MVFLNADNGILVQDL	PFEVKKKKFHIDFY	
					:		
20	orf88a	AKDFKPESILGASNLSFRGNVNI	SEGQSADVVF	LNADNGILVQDL	PFEVKKKKFHIDFY		
		210	220	230	240	250	260
	orf88.pep		40	50	60	70	80
			TGMPRDFASDIEV	TDKATGEKLE	RTIRVNHPLT	LHGITYIQAS	FADGSDLT
25	orf88a		TGMPRDFASDIEV	TDKATGEKLE	RTIRVNHPLT	LHGITYIQAS	FADGSDLT
			270	280	290	300	310
	orf88.pep		100	110	120	130	140
			ASREPVLKATSI	HQFPLEIGKH	KYRLEFDQFT	SMNVEDMSEGA	EEREKSLKST
30	orf88a		ASREPVLKATSI	HQFPLEIGKH	KYRLEFDQFT	SMNVEDMSEGA	EEREKSLKST
			330	340	350	360	370
	orf88.pep		160	170	180	190	200
			TQEGHKYTNXXXXX	YRIRDAPGQ	AVEYKNYML	PVLQEQQDYFW	ITGTRSLQQQ
35	orf88a		TQEGKKYTNIGPS	IVYRIRDAAG	QAVEYKNYML	PVLQEQQDYFW	ITGTRSGLQQ
			390	400	410	420	430
	orf88.pep		220	230	240	250	260
			PLDKQLKADTF	MALEFLKDG	EGRKRXVADAT	KGAPAEIREQ	FMLAAENTLNI
40	orf88a		PLDKQLKADTF	MALEFLKDG	EGRKRLVADAT	KGAPAEIREQ	FMLAAENTLNI
			450	460	470	480	490
	orf88.pep		280	290	300	310	320
			GLDEFITSNIPKE	QQDKMQGYFY	EMLYGMNAA	LDETXYTRYGL	PEWQQDEARN
45	orf88a		GLDEFITSNIPKE	QQDKMQGYFY	EMLYGMNAA	LDETIRRYGL	PEWQQDEARN
			510	520	530	540	550
	orf88.pep		340	350	360	370	
			DAYTGLTEYPAP	MLLQLDGFSEV	RSSGLQMTR	SXGPLLVL	
50	orf88a		DAYTGLTEYPAP	MLLQLDGFSEV	RSSGLQMTR	SPGALLVYLG	S VLLVLGTVLM
			570	580	590	600	610
	orf88a		AWVLFSDGKIRF	AMSSARSERDL	QKEFPKHVES	LQRLGKDLNH	DX
55			630	640	650	660	670

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

1 ATGAGTAAAT CCCGTAGATC TCCCCACTT CTTTCCCGTC CGTGGTTCGC
 51 TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
 101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT

151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
 201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTGTTATC ATGATGTTTT
 251 TGGTGGTTTC TACCAGTTTG TGCCTGATTC GCAATGTGCC GCCGTCTCTGG
 301 CGCGAAATGA AGTCTTTTCG GGAAGAGGTT AAAGAAAAAT CTCTGGCGGC
 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTGCCCA
 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
 451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAAATGGGG
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTTGCCGT GCGGGTGA
 551 TAGACAGTAA CCTGCTGTTG AAACGGGTA TGCTGACCGG TCGGATTGTT
 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
 701 AGAGTGCCTG TGTGGTTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
 751 GACTTGCCCT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
 801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG
 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG CCGGTTCCGA
 951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG
 1001 TGTGAAGGC AACATCCATA CACCAGTTTC CGTTGGAAAT TGGCAAACAC
 1051 AAATATCGTC TTGAGTTTCA TCAGTTTACT TCTATGAATG TGGAGGACAT
 1101 GAGCGAGGCG GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
 1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CCGCCCTTCC
 1201 ATTGTTTACC GTATCCGTGA TCGCGCAGGG CAGGCGGTCTG AATATAAAAA
 1251 CTATATGCTG CCGGTTTTCG AGGAACAGGA TTATTTTGG ATTACCGGCA
 1301 CCGCGACGGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
 1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAA GCGCACCTG
 1451 CCGAAATCCG CGAACAATC ATGCTGGCTG CGGAAACAC GCTGAACATC
 1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
 1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
 1651 CCCGAATGCG AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
 1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCCG GTTTGCAGAT GACCCGTTCC
 1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
 1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
 1901 ACGGCAAAAT CCGTTTGGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTGT
 1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRSPPL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
 51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
 101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVQ GFQKGTINRE
 151 DGSVLIAAKK GTMNKGWYIF AHVALIVICL GGLIDSNLLL KLGMLTGRIV
 201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
 45 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRNVHPLT
 301 LHGITIYQAS FADGGSDLTF KAWNLDGASR EPVVLKATSI HQFPLEIGKH
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKQYTNIGPS
 401 IVYRIRDAAG QAVEYKNYML PVLOEQDYFW ITGTRSGLQQ QYRWLRIPLD
 451 KQLKADTFMA LREFLKDEG RKRLVADATK GAPAEIREQF MLAAENTLNI
 50 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMAA LDETIRRYGL
 551 PEWQQDEARN RFLLSMDDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTS
 601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSDGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 orf88-1 MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 60 orf88a.pep QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88-1 QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88a.pep SSLLDVKIAPEVAKRYLEVQGFQKGTINREDGSVLIAAKKGTMNKGWYIFAHVALIVICL 180
 orf88-1 SSLLDVKIAPEVAKRYLEVQGFQKGTINREDGSVLIAAKKGTMNKGWYIFAHVALIVICL 180
 65 orf88a.pep GGLIDSNLLLKLGMMLTGRIVPDNQAVYAKDFKPEISILGASNLSFRGNVNISEGQSADVVF 240

	orf88-1	 GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSAADVVF	240
5	orf88a.pep	LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	 LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
10	orf88a.pep	LHGITIYQASFADGGSDLTFFKAWN LGDASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
	orf88-1	 LHGITIYQASFADGGSDLTFFKAWN LGDASREP VVLKATSIHQFPLEIGKHKYRLEFDQFT	360
15	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1	 SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	 PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
20	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	 GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSR	600
	orf88-1	 LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTSR	600
30	orf88a.pep	PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88-1	 PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD	672
35	orf88-1	 LQRLGKDLNHD	672

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

40	orf88.pep	MVFLNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	 MVFLNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
45	orf88.pep	PLTLHGITIYQASFADGGSDLTFFKAWN LGDASREP VVLKATSIHQFPLEIGKHKYRLEFD	120
	orf88ng	 PLTLHGITIYQASFADGGSDLTFFKAWN LGDASREP VVLKATSIHQFPLEIGKHKYRLEFD	120
50	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYRIRDAPGQAVEYKN	180
	orf88ng	 QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
55	orf88.pep	YMLPVLQEQQDYFWITGTRSLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRVAD	240
	orf88ng	 YMLPILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
60	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	 ATKDAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
65	orf88.pep	NAALDETXYTRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88ng	 NAALDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88.pep	TRSXGPELLVYL	371
	orf88ng	 TRSPGALLVYLGSVLLVLGTVFMFYVPKKRAWVLF SNXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

```

1  MVFLNADNGM LVQDLPFVEK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
5  51  KLERTIRVNH PLTLHGITIY QASFADGGS D LTFKAWNLRD ASREPVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLNDVRAV
151 TQEGKKYTN I GPSIVYRIRD AAGQAVEYKN YMLPILQDKD YFWLTGTRSG
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRLVAD ATKDAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKGQQDKMQG YFYEMLYGVM
301 NAALDETIIR YGLEPWQQDE ARNRFLLSHM DAYTGLTEYP APMLLQLDGF
10 351 SEVRSSGLQM TRSPGALLVY LGSVLLVLGT VFMFVVPKKR AWWLFSNXKI
401 RFAMSSARSE RDLQKEFPKH VESLQRLGKD LNHD*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

```

1  ATGAGTAAAT CCCGTATATC TCCCACACTT CTTTCCCGTC CGTGGTTCGC
15 51  TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
101 101 TGGCATCGGT TATCGGCACG GTGTTACAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGACC GTTTTGGACT CGGATTTTTG ATTTTGTGGG
201 TTTGTATGAT GTCTATGCTT CGGCATGGTT TGTCGTATAT ATGATGTTTC
251 TGGTGGTTTC TACCAAGTTT TGTTTAATCC GTAACGTTCC GCCGTTTGG
301 CGCGAAATGA AGTCTTTCCG GAAAAGGTT AAAGAAAAAT CTCTGGCGGC
20 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCCCCC GAAGTGGCCA
401 AACGTTATCT GGAGGTGCGG GGTTCAGG GAAAAACCGT CAGCCGTGAG
451 GACGGGTCCG TTCTGATTGC CGCAAAAAA GGCacaatga acaaATGGGG
501 CTATATCTTT GCcgaagtag ctTTGATTGT CATTGCGCTG GCGGGTTGA
551 TAGACAGTAA CTTGCTGCTG AAGCTGGGTA TGCTGGCCGG TCGGATTGTT
25 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AAAGTGCCTG TGTGGTTTTT CTGAATGCCG ACAACGGGAT GTTGGTTCAG
751 GACTTGCCTT TTGAAGTCAA ACTGAAAAA TTCCATATCG ATTTTACAA
801 TACCGGTATG CCGCGCGATT TTGCCAGCGA TATTGAAGTA ACGGACAAGG
30 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGCGAGT TTTGCCGACG GCGGTTCGGA
951 TTTGACATTC AAGGCGTGGA ATTTGAGGGA TGCTTCGCGC GAACCTGTCTG
1001 TGTGTGAAGC AACCTCCATA CACCAGTTTC CGTTGGAATC CGGCAAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
35 1101 GAGCGAGGGT GCGGAACGGG AAAAAAGCCT GAAATCCACT CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCCCTTCC
1201 ATCTGTATAC GCATCCGTGA TGcggCAGGG CAGGCGGTCTG AATATAAAAA
1251 CTATATGCTG CCGATTTTGC AGGACAAAGA TTATTTTGG CTGACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
40 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAAGA
1401 TGGGGAAGGG CGCAACGTC TGGTTGCCGA CGCAACCAA GACGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAATATC
1501 TTTGCGCAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
45 1551 CCCGAAAGGG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGCG AGCAGGATGA AGCGCGGAAC CGTTTCCTGC TGCACAGTAT
1701 GGATGCCTAT ACGGGGCTGA CGGAATATCC CGCGCCTATG CTGCTCCAGC
1751 TTGACGGGTT TTCCGAGGTG CGTTCCTCAG GTTTGCAGAT GACCCGTTCTG
50 1801 CCGGGTGCCT TTTTGGTCTA Tctcggtcg gtattgttgg TTTTGGgtac
1851 ggtaTttatg tTTTATGTGC GCGAAAAACG GGCGTGGgta tTGTTTTCag
1901 aCGGCAAAAT CCGTTTGTCT ATGtCTTcgg CCcgcagcga ACGGGATTTG
1951 cAGAaggaat TTCCAAAACA CGtcgAGAGC CTGCAACggc tcggcaaggA
2001 CttgaatCAT GACTga

```

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

```

55 1  MSKSRIPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLOQNQPQTD
51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVYSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVR GFQKTVSRE
151 DGSVLIAAKK GTMNKGYIF AQVALIVICL GGLIDSNLLL KLGMLAGRIV
60 201 PDNQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSADVVF LNADNGMLVQ
251 DLPFVEKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITIYQAS FADGGSDLTF KAWNLRDASR EPVVLKATSI HQFPLEIGHK
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PILQDKDYFW LTGTRSGLQQ QYRWLRIPLD
451 QQLKADTFMA LREFLKDGEK RKRLVADATK DAPAEIREQF MLAAENTLNI
65 501 FAQKGYLGLD EFITSNIPKG QQDKMQGYFY EMLYGVMNAA LDETIRRYGL

```


551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
 601 PGALLVYLGS VLLVLGTVFM FYVREKRAWV LFS DGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

```

5      orf88-1.pep  MSKSRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLVKFGSFWA 60
      orf88ng-1    MSKSRIPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQQNQPTDYLVKFGPFWT 60

10     orf88-1.pep  QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
      orf88ng-1    RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120

15     orf88-1.pep  SSSLDDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKGMTMNKGWYIFAHVALIVICL 180
      orf88ng-1    SSSLDDVKIAPEVAKRYLEVRFQGKTVSREDGSVLIAAKGMTMNKGWYIFAQVALIVICL 180

20     orf88-1.pep  GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240
      orf88ng-1    GGLIDSNLLKLGMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF 240

25     orf88-1.pep  LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLEERTIRVNHPLT 300
      orf88ng-1    LNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLEERTIRVNHPLT 300

30     orf88-1.pep  LHGITYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360
      orf88ng-1    LHGITYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360

35     orf88-1.pep  SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKTNI GPSIVYRIRDAAGQAVEYKNYML 420
      orf88ng-1    SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKTNI GPSIVYRIRDAAGQAVEYKNYML 420

40     orf88-1.pep  PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK 480
      orf88ng-1    PILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK 480

45     orf88-1.pep  GAPAEIREQFM LAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA 540
      orf88ng-1    DAPAEIREQFM LAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA 540

50     orf88-1.pep  LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPM LLQLDGFSEVRSSGLQMTRS 600
      orf88ng-1    LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPM LLQLDGFSEVRSSGLQMTRS 600

55     orf88-1.pep  PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES 660
      orf88ng-1    PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES 660

60     orf88-1.pep  LQRLGKDLNHD 671
      orf88ng-1    LQRLGKDLNHD 671
  
```

Furthermore, ORF88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```

55     gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
      Score = 94.4 bits (231), Expect = 2e-18
      Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

      Query: 16  FAFSSMRFAVALLSLLGIASVIG-TVLQQNQPTDYLVKFGPFWTRIFDFLGLYDVYAS 74
      + F +S++ A+ ++ +LGI S++G T ++QNQ YL +FG L L DV+ S
      Sbjct: 80  YDFLASLKLAI FIMLVLGILSMLGSTYIKQNSFEWYLDQFGYDVGIWIWKLWLNDFVHS 139

60     Query: 75  AWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK 134
      +++++ ++ L V+ C I+ +P W++ S +E++ + A +H + VKI P+ K
      Sbjct: 140  WYYILFIVLLAVNLIFCSIKRLPRVWQAFS-KERILKLDEHAEKHLKPITVKI-PDKDK 197

65     Query: 135 --RYLEVRFQGKTVSREDGSVLIAAKGMTMNKGWYIFAQVALIVICLGGGLIDSNLLKL 192
      ++L +GF+ V E ++ A+KG ++ G +AL+VI G LID
      Sbjct: 198  VLKFLKKGFK-VFVEEGNKLYVFAEKGRFSRLGVYITHIALLVIMAGALID----- 249
  
```

Example 40

1	ATGATGAGTA	ATAmAATGGm	ACAAAAGGG	TTTACATTGA	TTGmGmTGAT
51	GATAGTCGTC	GCGTATCATCG	GCATTATCAG	CGTCATTGCC	ATACCTTCTT
101	ATCmAAGTATA	TATTGAAAAA	GGCTATCAGT	CCAGCTTTA	TACGGAGATG
151	GyCGGTATCA	ACAAATATTTT	CAAACAGTTT	ATTTTGAAAA	ATCCCCTGGA
201	CGATAATCAG	ACCATCGAGA	ACAAACTGGA	AATATTTGTC	TCAGGCTATA
251	AGATGAATCC	GAAAATTTGCC	AAAAAaTATA	GTGTTTCGGT	AAAGTTTGTC
301	GATAAGGAAA	AATCAAGGGC	ATACAGGTTG	TCGGCGGTC	GCAAGCGCGG
351	GACGGGTATT	ACTTTGTCCG	TATGGATGAA	CACGGTGGGC	GACGGATACA
401	AATGCCCTTA	TGCCGCTTCT	GCCCAAGCCC	ATTTGGAGAC	CTTGTCTCTA
451	GATGTCGGCT	GTGAAGCCTT	CTCTAATCGT	AAAAAATAA	

```

1  MMSNXMXQKG FTLIXXMIVV AILGIISVIA IPSYXSIEK GYQSQLYTEM
51 XGINNISQKF ILKNPLDDNQ TIENKLEIFV SGYKMNP KIA KYKSVSVK FV
101 DKEGSRAYLR VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLSS
151 DVGCEAFSNR KK*

```

1	ATGATGAGTA	ATAAAATGGA	ACAAAAGGG	TTTACATTGA	TTGAGATGAT
51	GATAGTCGTC	GCGATATACG	GCATTATCAG	CGTCATTGCC	ATACCTTCTT
101	ATCAAAGTTA	TATTGAAAAA	GGTATCAGT	CCGACGTTTA	TACGGAGATG
151	TCGCGTATCA	ACAATATTTT	CAAAACAGTT	ATTTTGAAAA	ATCCCCTGGA
201	CGATAATCAG	ACCATCGAGA	ACAAATGGA	AATATTTGTC	TCAGGCTATA
251	AGATGAATCC	GAAAAATGCC	AAAAAATATA	GTGTTTCGGT	AAAGTTTGTC
301	GATAAGGAAA	AATCAAGGGC	ATACAGTTTG	TCGGCGGTC	GCAAGCGCGG
351	GACGGGTAT	ACTTTGTCCG	TATGATGAA	CACGGTGAGC	GACGGATACA
401	AATGCGGTGA	TGCCGCTTCT	GCCCAAGCC	ATTTGGAGC	CTTGTCTCTA
451	GATGTCGGCT	GTGAAGCCTT	CTCTAATCGT	AAAAAATAA	

```

1  MMSNKMEQKG  FTLIEMMIVV  AILGIISVIA  IPSYQSYIEK  GYQSQLYTEM
51  VGINNISKQF  ILKNPLDDNQ  TIENKLEIFV  SGYKMNPKIA  KKYSVSVKfV
101 DKEKSRAYRL  VGVPKAGTGY  TLSVWMNSVG  DGYKCRDAAS  AQAHLETLS
151 DVGCEAFSNR  KK*

```

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:

```

orf89 8  QKGFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQFILKNPL- 66
        QKGFTLI  MIV+AI+GI++ +A+P+Y  Y  +  S+      G  +  ++ L  +  +
Pile   5  QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGQKSAVTEYYLNHGIW 64

5  orf89 67 -DDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGYTLVSW 125
        DN  +      +G  +  KI  KY  SV      +      GV  K  G  LS+W
Pile   65  PKDNTS-----AGVASSDKIKGKYVQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW 115

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

```

15 orf89.pep  MMSNXMQKQKFTLIXXMIVVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQF
        10      20      30      40      50      60
orf89a      MMSNKMEQKQKFTLIXXXXXAIXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX
        10      20      30      40      50      60

20 orf89.pep  ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGY
        70      80      90      100     110     120
orf89a      ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHVFVNEEKPRAYSLVGVPKTGTGY
        70      80      90      100     110     120

25 orf89.pep  TLSVWMNSVGDGYKCRDAASQAHALETLSDDVGCEAFSNRKKX
        130     140     150     160
orf89a      TLSVWMNSVGDGYKCRDAASARAHLETLSDDVGCEAFSNRKKX
        130     140     150     160

```

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

```

30 1 ATGATGAGTA ATAAAATGGA ACAAAGAGG TTTACATTGA TTGNGANGNT
51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
101 ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTT CAAACAGTNT ATTTTGAAAA ATCCCCTGGA
201 CGATAATCAG ACCATCAAGA GCAAACAGT AATATTGTG TCAGGCTATA
35 251 AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTTGTC
301 AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGG GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGCTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

40 This encodes a protein having amino acid sequence <SEQ ID 342>:

```

1 MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXSYIEK GYQSOLYTEM
51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHVF
101 NEEKPRAYSL VGVPKTGTGY TLSVWMNSVG DGYKCRDAAS ARAHLETLS
151 DVGCEAFSNR KK*

```

45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

```

50 orf89a.pep  MMSNKMEQKQKFTLIXXXXXAIXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX
        10      20      30      40      50      60
orf89-1      MMSNKMEQKQKFTLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF
        10      20      30      40      50      60

55 orf89a.pep  ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHVFVNEEKPRAYSLVGVPKTGTGY
        70      80      90      100     110     120
orf89-1      ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAVRLVGVPKAGTGY
        70      80      90      100     110     120

60 orf89a.pep  TLSVWMNSVGDGYKCRDAASARAHLETLSDDVGCEAFSNRKKX
        130     140     150     160
orf89-1      TLSVWMNSVGDGYKCRDAASQAHALETLSDDVGCEAFSNRKKX

```

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

orf89	MMSNXMXQKGETLIXXMIVVAILGIISVIAIPSYXSYIEKGYSQLYTEMGXINNISKQF : 60
orf89ng	MMSNKMEQKGTFLIEMMIVVTILGIISVIAIPSYQSYYIEKGYSQLYTEMGVINNVLKQF 60
orf89	ILKNPLDDNQTIENKLEIFVSGYKMNPKIACKYSVSVKFVDKEKSRA YRLVGVPKAGTGY : : : : : : 120
orf89ng	ILKNPQQDDNDTLKSKLKIFVSGYKMNPKIACKYSVSVRFVDAEKPRAYRLVGVPNAGTGY 120
orf89	TLSVWMNSVGDGKYCRDASAQA HLETSSDVGC EAFSNRKK 162 : : : 162
orf89ng	TLSVWMNSVGDGKYCRDATSA QAYSDTLSADSGCEAFSNRKK 162 162

	1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTGAGATGAT
20	51	GATAGTTGTC	ACGATACTCG	GCATCATCAG	CGTCATTGCC	ATACCTTCTT
	101	ATCAGAGTTA	TATTGAAAAA	GGCTATCAGT	CCCAGCTTTA	TACGGAGATG
	151	GTCGGGTATCA	CAATGTGTTCT	CAAAAGCTTTT	ATTTTGAAAA	ATCCCCAGGA
	201	CGATAATGAT	ACCCTCAAGA	GCAAACTGAA	AATATTGTGC	TCAGGCTATA
	251	AGATGAATCC	AAAAAttgCC	AAAAAATATA	GTGTTTCGGT	aaggttttGTC
25	301	gatCCGGA	GAACAAGGCG	ATACAGGTTG	GTCCGGCTTC	CGAACCGGGG
	351	GACGGGTTAT	ACTTTGTCGG	TATGGATGAA	CAGCGTGGGC	GACGGATACA
	401	AATGCCGTGA	TGCCACTTCT	GCCCAGGCCT	ATTCCGGACAC	CTTGTCCGCA
	451	GATAGCGGCT	GTGAAGCTTT	CTCTAATCGT	AAAAAATAG	

30 1 MMSNKMEOQKG FTLIEMMIVV TILGIISVIA IPSYQSYIEK GYQSOLYTEM
51 VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
101 DAEKPAYRL VGVPNAGTGY TLSVWMNSVG DGYKCRDATS. AQAYSDTLSA
151 DSGCEAFSNR KK*

35 identity in 162 aa overlap:

		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGFTLIEMMIVVAILGIIISVIAIPSYQS	YIEKGYQS	QLYTEMVGINNISKQF			
40	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIIISVIAIPSYQS	YIEKGYQS	QLYTEMVGINNVLKQF			
		10	20	30	40	50	60
	orf89-1.pep	ILKNPLDDNQTIENKLEIFVSGYKMNP	KI AKKYSVSVKFVDKEKS	RAYRLVGVPKAGTGY			
			: : :				
45	orf89ng	ILKNPQDDNDTLKSKLKI FVSGYKMNP	KI AKKYSVSVRFVDAEKPRAYRLVGVPNAGTGY				
		70	80	90	100	110	120
	orf89-1.pep	TLSVWMNSVGDGYKCRDAASAQA	HLETLSSDVGCEAFS	NRRKKX			
				: :	:		
50	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYS	DTLSADSGCEAFS	NRRKKX			
		130	140	150	160		

55 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAAGCGT GGCCn.AGAA TTTCAACCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

15 1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
    51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

20 1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
    51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
    101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
    151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
    201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    251 AACAAAGCGT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
    301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
    351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
    401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
    451 GTTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
    551 GACTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAATA A
```

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKLNAN VNVKDNPIVN KGGKEIIVRA EVGVPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGASLVTYR NQFGEIIRAK GVDGLIAELK AKNGGK*
```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

```

40  orf91.pep      10      20      30      40      50      60
    MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
    |||||:|||||
    orf91a        MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
                  10      20      30      40      50      60

45  orf91.pep      70      80      90
    YDFQRM TALAVGNPWXTXSDXQKQALAXEFQP
    |||||
    orf91a        YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLNANVNVKDNPIVN
                  70      80      90      100     110     120
```

orf91a KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51 CGGCATGGCA TTTGCCGCC CTGCCGACGC GGTAACCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
     151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTGATT TCCAACGTAT
     201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
    10 251 AACAAAGCGTT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301 GGCACGATGC TGAATTTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
     401 TACCGGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
    15 501 CGTGATACCG AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
     551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
     51 RQKAEAYAIP YFDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
    20 101 GTMLKLKNAN VNVKDNPIVN KGGKEIIVRA EVGVPQKPV NMDFTTYQSG
     151 GKRYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGSK*

```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

      10      20      30      40      50      60
25  orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91-1 MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
      10      20      30      40      50      60

      70      80      90     100     110     120
30  orf91a.pep YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLKNANVNVKDNPIVN
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91-1 YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLKNANVNVKDNPIVN
      70      80      90     100     110     120

      130     140     150     160     170     180
35  orf91a.pep KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91-1 KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
      130     140     150     160     170     180

      190
40  orf91a.pep GVDGLIAELKAKNGSKX
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91-1 GVDGLIAELKAKNGGKX
      190
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

50  orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP 60
      :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91ng VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP 60

      YFDFQRM TALAVGNPWXTXSDXQKQALAXEFQP 93
55  orf91.pep |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  orf91ng YFDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKFNATVNVKDNPIVN 120

```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

-230-

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTYVR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTTGCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTGTACC ATCCTCAAAA GCGGCGACGC GGCTTCTGCA
10 GACCCAAAAG CCGAAGCCTA TCGGTTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAAA
251 AACAAAGCGT GGCCAAAGAA TTCAAACCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCGCGTCA GAAGCCCGTC AATATGGACT TTACCACCTA CCAAAGCGGC
15 GGCAAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
501 CGGTGTACCG AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRQNATQVLT ILKSGDAASA
20 51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTYVR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25 orf91-1.pep      10      20      30      40      50      60
    MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
orf91ng-1      10      20      30      40      50      60
    MKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP

30 orf91-1.pep      70      80      90      100     110     120
    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKLKNANVNVKDNPIVN
orf91ng-1      70      80      90      100     110     120
    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPIVN

35 orf91-1.pep      130     140     150     160     170     180
    KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEI IKAK
orf91ng-1      130     140     150     160     170     180
    KGGKEIVVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTYVRNQFGEI IKAK

40 orf91-1.pep      190
    GVDGLIAELKAKNGGKX
orf91ng-1      190
    GIDGLIAELKAKNGGKX

```

In addition, ORF91ng-1 shows homology to a hypothetical *E.coli* protein:

```

50 sp|P45390|YRBC_ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
    REGION PRECURSOR (F211) >gi|606130 (U18997) ORF f211 [Escherichia coli]
    >gi|1789583 (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic
    region [Escherichia coli]Length = 211

```

Score = 70.6 bits (170), Expect = 6e-12

Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)

```

55 Query: 59 VPYDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGTMLKFKNATVNVKDNPI 118
    +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
    Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122

```

```

60 Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTYVRNQF 174
    G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
    Sbjct: 123 QPLGDKTIVPIRVTIIDPNGRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

Query: 175 EIIKAKGIDGLIAELKA 191
 +++ KGIDGL A+LK+
 Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal
 5 protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes,
 could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

10      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
     101  TGATCAGCGA TACCCTCATC TCAAATACA GTTTTGGnnn nnnnnnnnnn
     151  nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTCCGACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AAACGATAAC AAAAACCCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

20      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
      51  XXXXAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

25      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTC AAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
     101  TGACACGCGA TACCCTCACC TCAAATACA GTTTTGACGA AACCGTCAGC
     151  CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TGGCCGTCAT
     201  CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTCCGACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451  AAACGATAAC AAAAACCCGT AGGCGAATAA
  
```

35 This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

```

      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTTHLT SKYSFDETVS
      51  RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVGE*
  
```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.meningitidis*:

```

45      orf97.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXXXXAIKSKG
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf97a          MXHILPLXASALCISTASXHPASEPQTQNETAMTTHLTISKYSFDETVSRLETAIKSKG
                        10      20      30      40      50      60
  
```


-232-

		70	80	90	100	110	120
orf97.pep		MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
orf97a		MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
5		70	80	90	100	110	120
		130	140	150	160		
orf97.pep		VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
10	orf97a	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

	1	ATGANACACA	TACTCCCCCT	GANTGNCGCA	TCCGCACTCT	GCATTTC AAC
	51	CGCTTCGGNN	CATCCTGCCA	GCGAACCGCA	AACCCAAAAC	GAAACCGCTA
15	101	TGACCACGCA	TACCCTCACC	TCAAATACA	GTTTGTACGA	AACCGTCAGC
	151	CGCCTTGAAA	CGCCATAAA	AAGCAAAGGG	ATGGACATT	TGCGCGTCAT
	201	CGACCATCAG	GAAGCCGCCC	GCCGAAACGG	CTTAACGATG	CAGCCGGCAA
	251	AAGTCATCGT	CTTCGGCACG	CCCAAAGCCG	GTACGCCGCT	GATGGTCAAA
	301	GACCCCGCCT	TCGCCCTGCA	ACTGCCCTG	CGCGTCNTCG	TTACCGAAAC
20	351	GGACGGCAAA	GTACGCGCCG	CCTATACCGA	TACGCGCGCC	CTCATCGCCG
	401	GCAGCCGCAT	CGGTTTCGAC	GAAGTGGCAA	ACACTTTGGC	AAACGCCGAA
	451	AAACTGATAC	AAAAAACCAT	AGGCGAATAA		

This encodes a protein having amino acid sequence <SEQ ID 360>:

	1	MXHILPLXXA	SALCISTASX	HPASEPQTQN	ETAMTHTLT	SKYSFDETVS
25	51	RLETAIKSKG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVK
	101	DPAFALQLPL	RVXVTETDGK	VRAAYTDTRA	LIAGSRIGFD	EVANTLANAE
	151	KLIQKTIGE*				

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

		10	20	30	40	50	60
30	orf97a.pep	MXHILPLXXASALCISTASXHPASEPQTQN	ETAMTHTLT	SKYSFDETVS	RLETAIKSKG		
	orf97-1	MXHILPLIAASALCISTASAHASEPSTQN	ETAMTHTLT	SKYSFDETVS	RLETAIKSKG		
		10	20	30	40	50	60
35	orf97a.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVXVTETDGK					
	orf97-1	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK					
		70	80	90	100	110	120
40	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		
45	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX					
		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N. gonorrhoeae*:

50	orf97.pep	MXHILPLIAASALCISTASAHASEPSTQN	ETAMITHTLISKYSFGXXXXXXXXXAIKSKG	60
	orf97ng	MXHILPPIAASAFICISTASAHAPAGKPPTQN	ETAMTHTLTISKYSFDETVSRLETAIKSKG	60
55	orf97.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK		120
	orf97ng	MDIFAVIDHQEARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK		120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTIGEX		159
60	orf97ng	VRTAYTDTRALIVSRISFDEVANTLANAEKLIQKTIGEX		159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

1  MKHILPPIAA  SAFCISTASA  HPAGKPPTQN  ETAMTTHTLT  SKYSFDETVS
5  51  RLETAIKSKG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
101 DPAFALQLPL  RVLVTETDGK  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

1  ATGAAACACA  TACTCCCcct  gatcgccgca  TccgcactCT  GCATTTC AAC
10  51  CGCTTCGGCA  CACCCTGCCG  GCAAACCGCC  CACCCAAAAC  GAAACCGCTA
101 TGACCACGCA  CACCCTCACC  TCGAAATACA  GTTTTGACGA  AACCCTCAGC
151 CGCCTTGAAA  CCGCCATAAA  AAGCAAAGGG  ATGGACATTT  TTGCCGT CAT
201 CGACCATCAG  GAAGCGGCAC  GCCGAAACGG  CCTGACCATG  CAGCCGGCAA
251 AAGTCATCGT  CTTCGGCACG  CCCAAGGCCG  GTACGCCgct  GATGGTCAAA
301 GACCCCGCCT  TCGCCCTGCA  ACTGCCCTG  CGCGTCCTCG  TTACCGAAAC
15  351 GGACGGCAAA  GTACGCACCG  CCTATACCGA  TACGCGCGCC  CTCATCGTCG
401 GCAGCCGCAT  CAGTTTCGAC  GAAGTGCGAA  AACTTTGGC  AAACGCCGAA
451 AAAGTATAC  AAAAAACCGT  AGGCGAATAA
```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

1  MKHILPLIAA  SALCISTASA  HPAGKPPTQN  ETAMTTHTLT  SKYSFDETVS
20  51  RLETAIKSKG  MDIFAVIDHQ  EAARRNGLTM  QPAKVIVFGT  PKAGTPLMVK
101 DPAFALQLPL  RVLVTETDGK  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

```

25      10      20      30      40      50      60
orf97-1.pep  MKHILPLIAASALCISTASAHASEPSTONETAMTTHTLTSKYSFDETVSRLETAIKSKG
orf97ng-1    MKHILPLIAASALCISTASAHASEPSTONETAMTTHTLTSKYSFDETVSRLETAIKSKG
      10      20      30      40      50      60
30      70      80      90     100     110     120
orf97-1.pep  MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
orf97ng-1    MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
      70      80      90     100     110     120
35      130     140     150     160
orf97-1.pep  VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
orf97ng-1    VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX
40      130     140     150     160
```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

15 Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is
87.9% over the same 199 aa overlap.

	1	ATGGCTTTTA	TTACGCGCTT	ATTCAAAGC	ATTAAACAAT	GGCTTGTGCT
20	51	GCTGCCGATG	CTTTCCGTTT	TGCCGGACGC	GGCGGCGGAG	GGGATAGATG
	101	TGAGCCGCGC	CGAAGCGAGG	ATAANCGACG	GC GGCGACGT	TTCCATTNAGN
	151	AGCCCGTTCC	AAACCGAGCT	GCCCGACCAG	CTCAANNNG	CGNNGNGCCG
	201	GGGCGTGNCG	CTCAACTNTA	CCTTAAGNTG	GCAGCTTTCC	GCCCGGATAA
	251	TCGCTTCTTA	TCGGTTTNA	TTGGGGCAAC	TGATTGGCGA	TGACACGNAT
25	301	ATTGACTACA	AACTGAGTTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
	351	CGTCGCGCGC	TTTTCGACAG	ANTACGCAC	CTTGGATGCG	GCATTGCGCG
	401	CGACCGGCGC	GGTTGCCAAC	TGGAAAGTCC	TGAACAAAGG	CGCGCTGTCC
	451	GGTGCCGGA	CAGGGGAAAC	CAAGGCGGAA	ATCCGCTGTA	CGCTGTCCAC
	501	TTCAAAACTG	CCCAAGCCTT	TTCAAACTAA	TGCATTGACT	TCTCAAACT
	551	GGCATTTGGA	TTCGGTTTGG	AAACCTCTAA	GCATCATCGG	GAACAATAAA

1 MAFITRLFKS IKQVLVLLPM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX
51 SRFQTELPDQ LQXAXXRGVX LNXTLXWQLS APTIASYRFK LQQLIGDDDX
101 IDYKLSFHPL TNRYRVTVGA FSTXYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFOINALT SQNWHLDSGW KPLNIIGNK*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N. gonorrhoeae*:

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAAACAAT GGCTTGTGCT
51  GTTGCCGATA CTCTCCGTTT TGCCGGACGC GGCGGCGGAG GGCATGCGCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
5  151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCC GCGCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAAT
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCCTG ACCGCGTTAC
351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATGCGCGG
10 401 CGACCGGCGC GGTGCGCAAC TGGAAAGTCC TGAACAAAGG CGCGTGTGCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCCTGA CGCTGTCCAC
501 TTCAAACTG CCCAAGCCTT TCCAATCAA CGCATTGACT TCTCAAACT
551 GGCATTTGGA TTCGGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15 1  MAFITRLFKS IKQVLVLLPI LSVLPDAAAE GIAATRAEAR ITDGRLSIS
51  SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 IDYKLSFHPL TNRYRVTGVA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFQINALT SQNWHLD SGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTgACGGTG
151 TCGGTGTGTG GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
35 201 CACCGCCGAC AAAGACAcCT TGTTCAAAAC CCTGTPCCTG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC gCCGCCGGCa TCGGGCTGGT
351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTC
401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcAAG
45 451 CTCGCCATCC TGCTGCTG.T GCCGCTGACG GTCGGGCTGC TGCATTTC
501 AGCGAACACC GCCGTCCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACCGACCGT TTTCCGCCGC CGTCCTGCAC CGGGG.TGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
45 701 GTTTGTTCCT GAAAAAATAT GCCCGCTGG AACAGCTCGG CGTTTATTTC
751 ATGGGTATTT CGTTCCGCGG GGCGGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
901 GCCCTCTGC. TGACCGGCAT TTTCTCGCCC CTTGCCCTCC TCCTGCTGCC
50 951 GGAAACTAC GCCGCCGTCC GGTTTATCGT CGTATCGTGT ATG.TGCCGC

```

-237-

5
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTCGTTCTCAAGACC
1201 GAAAGCTCyt GCCGCCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGGCAAA CTATCCCCTG TTTGCCGGCG TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GGTTCCTCAT TATGA

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

15
20
25
30
35
40
45
1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLMEGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGXYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MXPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPAV PXGAAVACAA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAAYLA
20 451 GCILRHRKDL HKLFHYLKKQ GFPL*

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

25
30
35
40
45
50
1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCCCGCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGCGCGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACGCGCGAC AAAGACACCT TGTTCAAAAC CCTGTTCCTG CCGCCGCTGC
251 TGCTGCGCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCACT CGACGATGCC GCCGCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCATTTCCTC
501 AGCGAACACC GCCGTCCTGA CCGCGGTTTA CGCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGACCCGT TTTGCGCCGC CGTCTGACG CCGGGGCTGC GCTACGGCAT
35 651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTGTTCCTT GAAAAATAT GCCGCGCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GCGCGCATT TGTTCCTCAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCCCGCCT CTGCGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
40 901 GCCCTCTGCC TGACCGGCAT TTTCTGCCC CTTGCCCTCC TCCTGCTGCC
951 GGAAACTAC GCCGCCGTCC GGTTCATCGT CGTATCGTGT ATGCTGCCGC
1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG
45 1151 CGGTGCTGCTG TGCCGCCTCA TTCTGGCTGT TTTTTCCTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCTGCCGC TTTATCTGCA
1251 CACATTGTTT TGCTGACCT CCTCGGCGGC CTACACCTGC TTCGGCACGC
1301 CGGCAACTA TCCCTGTTT GCCGCGTAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGGAA AGATTGAC AAATGTTTC ATTATTGAA
50 1401 AAAACAAGT TTCCATTAT GA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

55
60
1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSFLPIRFL LVLMEGRAL AFSSAQLVPK
151 LAILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSFSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAFKTE
60 401 SSCRLWQPLK RLPLYLHTLF CLTSSAAYTC FGTPANYPLF AGVWAAAYLAG
451 CILRHRKDLH KLFHYLKKQ FPL*

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- 5 ORF10 shows homology with the epsM gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

10 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLLQSIFSTVW 270
L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W
Sbjct: 210 LYYALPLIPSSILWLLNASSRYFVLFLLGAGANGLLAVATKIPSIISIFNTIFTQAW 267

15 Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
L + G++GS +L +++PL ++ + G L QT A L + ++ + + A +R
Sbjct: 12 LVFTIGNLGSKLLVFLVPLTYTYAMTPQEYGMADLYQTTANLLLPLITMNVFDATLR 68

20 Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLL PENYA AAVRFTVVSCMLPFLFYTLTEISGIGLNVVRKTRPIXXXXXXXXXX 366
+ P+ ++ +YA+ V ML LF + ++ G ++T+ +
Sbjct: 305 VLKPIVEKVSSDYASSWQYVFFMLSMLFSSFSDFFGTNYIAAKQTKGVFMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N. meningitidis*:

30	orfl0.pep	10	20	30	40	50	60
		MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	orfl0a	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
35	orfl0.pep	70	80	90	100	110	120
		YVREYYATADKDTLFTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
	orfl0a	YVREYYAAADKDTLFTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
40	orfl0.pep	70	80	90	100	110	120
		YVREYYATADKDTLFTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
	orfl0a	YVREYYAAADKDTLFTLFLPPLLSAAAIAALLSRPSLPSEILFSLDDAAAGIGLVLE					
45	orfl0.pep	130	140	150	160	170	180
		LSFLPIRFLLLVLRMEGRALAFSSAQLVLPKLAILLXPLTVGLLHFPANTAVLTAVYALA					
	orfl0a	LSFLPIRFLLLVLRMEGRALAFSSAQLVSKLAILLXPLTVGLLHFPANTAVLTAVYALA					
50	orfl0.pep	190	200	210	220	230	240
		NLAAAFLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAIWGLASADRLFLKKY					
	orfl0a	NLAAAFLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAIWGLASADRLFLKKY					
55	orfl0.pep	250	260	270	280	290	300
		AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orfl0a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS					

-239-

		310	320	330	340	350	360
	orf10.pep	ALCXTGIFSP	LASLLLPENYA	AVRFIVVSC	MPPLFCTLAE	ISGIGLNVV	RKTRPIALAT
5	orf10a	ALCLTGIFSP	LASLLLPENYA	AVRFIVVSC	MLPPLFCTL	VEISGIGLNV	VVRKTRPIALAT
		310	320	330	340	350	360
		370	380	390	400	410	419
	orf10.pep	LGALAANLLL	LGLDRAVPA	R-PXGA	AVACAASF	WLFFAFKTE	SSCRLWQPLKRLPLYLHT
10	orf10a	LGALAANLLL	LGL--AVPS	GGGARGA	AVACAASF	WLFVFKTE	SSCRLWQPLKRLPLYMHT
		370	380	390	400	410	
		420	430	440	450	460	470
	orf10.pep	LFCLTSSA	AYTCFGT	PANYPLF	AGVWAA	YLAGCILR	HRKDLHKLFHYLKKQGFPLX
15	orf10a	LFCLASSA	AYTCFGT	PANYPLF	AGVWAV	YLAGCILR	HRKDLHKLFHYLKKQGFPLX
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
20	51	GGTTTTAGCC	GTCATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCCTGCCCG
	101	ACGACATCGG	ACGCATCGTG	CTGATGCAGA	CGGCGCGGG	GCTGACGGTG
	151	TGGTGTGTGT	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCTCTG	CCGCCGCTGC
	251	TGTCTGCCCG	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	ATCCCTGCCCG
25	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCCGGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTCTCTTA	CTGGTTTTCG
	401	GTATGGAAGG	ACGCGCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGTCCAAG
	451	CTCGCCATCC	TGCTGCTGCT	GCCGCTGACG	GTCGGGCTGC	TGCACTTTCC
	501	GGCGAACACC	GCCGCTCTGA	CCGCCGTTTA	CGCGCTGGCA	AACCTTGCCG
30	551	CCGCCGCCTT	TTTGCTGTTT	CAAAACCGAT	GCCGCTGAA	GGCCGTCGGG
	601	CGCGCACCGT	TTTCATCCGC	CGTCTGCAT	CGCGGCTGCG	GCTACGGCAT
	651	ACCGATCGCA	CTAAGCAGCA	TCGCCTATTG	GGGGCTGGCA	TCCGCCGACC
	701	GTTTGTTCCT	GAAAAAATAT	GCCGGCCTAG	AACAGCTCGG	CGTTTATTCG
	751	ATGGGTATTT	CGTTCGGCGG	AGCGGCATTA	TTGTTCCAAA	GCATCTTTTC
35	801	AACGGTCTGG	ACACCTATA	TTTTCCGCGC	AATCGAAGCA	AACGCCCCGC
	851	CCGCCCGCCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCTCTC
	901	GCCCTCTGCC	TGACCGGCAT	TTTCTCGCCC	CTCGCCTCCC	TCCTGCTGCC
	951	GGAAACTAC	GCCGCCGTCC	GGTTTATCGT	CGTATCGTGT	ATGCTGCCTC
	1001	CGCTGTTTTG	CACGCTGGTA	GAAATCAGCG	GCATCGGTTT	GAACGTGCTC
40	1051	CGAAAAACAC	GCCCGATCGC	GCTCGCCACC	TGGGCGCGC	TGGCGGCAAA
	1101	CTGCTGCTG	CTGGGCTTG	CCGTACCGTC	CGGCGGCGCG	CGCGGCGCGG
	1151	CGGTTGCCTG	TGCCGCCTCA	TTTTGGCTGT	TTTTTGT	CAAGACCGAA
	1201	AGCTCTGCTC	GCCTGTGGCA	GCCGCTCAAA	CGCCTGCCGC	TTTATATGCA
	1251	CACATGTGTC	TGCCTGGCCT	CCTCGGCGGC	CTACACCTGC	TTCGGCACTC
45	1301	CGGCAAACTA	CCCCCTGTTT	GCCGGCGTAT	GGGCGGTATA	TCTGGCAGGC
	1351	TGCATCTGCG	GCCACCGGAA	AGATTGTCAC	AAACTGTTTC	ATTATTTGAA
	1401	AAAACAAGGT	TTCCATTAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 378>:

	1	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV
50	51	SVLCLGLDQA	YVREYYAAAD	KDTLFKTLFL	PLLSAAAIA	ALLLSRPSLP
	101	SEILFSLDDA	AAGIGLVLFE	LSFLPIRELL	LVLRMGRAL	AFSSAQLVSK
	151	LAILLPLPLT	VGLLHFPANT	AVLTAVYALA	NLAAAFLFF	QNRCLKAVR
	201	RAPFSSAVLH	RGLRYGIPIA	LSSIAYWGLA	SADRLFLKKY	AGLEQLGVYS
	251	MGISFGGAAL	LFQSIFFSTVW	TPYIFRAIEA	NAPPARLSAT	AESAALLAS
55	301	ALCLTGIFSP	LASLLLPENY	AAVRFIVVSC	MLPPLFCTLV	EISGIGLNVV
	351	RKTRPIALAT	LGALAANLLL	LGLAVPSGGA	RGAAVACAAS	FWLFFVFKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAYTC	FGTPANYPLF	AGVWAVYLAG
	451	CILRHRKDLH	KLFHYLKKQG	FPL*		

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

		10	20	30	40	50	60
60	orf10-1.pep	MDTKEILXYA	AGSIGSAVL	AVIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
	orf10a	MDTKEILGYA	AGSIGSAVL	AVIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
		10	20	30	40	50	60
65							

-240-

		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTLFKTLFPLP	LLSAAAIAALLSRP	SLPSEILFSLDDAAAGIGLVLFE			
	orf10a	YVREYYAAADKDTLFKTLFPLP	LLSAAAIAALLSRP	SLPSEILFSLDDAAAGIGLVLFE			
5		70	80	90	100	110	120
	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQ	LVPKLAILLXPLTVGLLHFPANTAVLTAVYALA				
	orf10a	LSFLPIRFLLLVLRMEGRALAFSSAQ	LVSKLAILLXPLTVGLLHFPANTAVLTAVYALA				
10		130	140	150	160	170	180
	orf10-1.pep	NLAAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSS	IAYWGLASADRLFLKKY				
	orf10a	NLAAAAFLLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSS	IAYWGLASADRLFLKKY				
15		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYSMGISFGGAALLFQSI	FSTVWTPYIFRAIEENAPPARLSATAESAAALLAS				
	orf10a	AGLEQLGVYSMGISFGGAALLFQSI	FSTVWTPYIFRAIEANAPPARLSATAESAAALLAS				
20		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSPLASLLLPENYA	AVRFIVVSCMXPPLFCTLA	ISGIGLNVVRKTRPIALAT			
	orf10a	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFCTLVE	ISGIGLNVVRKTRPIALAT			
25		310	320	330	340	350	360
	orf10-1.pep	LGALAANLLLLGLDRAVPAR-PXGA	AVACAASF	WLF	FAFKTESSCRLWQPLKRLPLYLHT		
	orf10a	LGALAANLLLLGL--AVPSGG	ARGA	AVACAASF	WLF	FFVFKTESSCRLWQPLKRLPLYMHT	
30		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSAAYTCFGTPANYPLFAGV	WAAAYLAGCILRHRKDLHKL	FHYLKKQGFPLX			
	orf10a	LFCLASSAAYTCFGTPANYPLFAGV	WAVYLAGCILRHRKDLHKL	FHYLKKQGFPLX			
35		420	430	440	450	460	470
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA				
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA				
40		420	430	440	450	460	470

Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

45 *gonorrhoeae*:

	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILP	LLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
50	orf10ng.pep	YVREYYAAADKDTLFKTLFPLP	LLSAAAIAALLSRP	SLPSEILFSLDDAAAGIGLVLFE 120
	orf10nm	YVREYYATADKDTLFKTLFPLP	LLSAAAIAALLSRP	SLPSEILFSLDDAAAGIGLVLFE 120
55	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQ	LVPKLAILLXPLTVGLLHFPANTS	SVLTAVYALA 180
	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQ	LVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	180
	orf10ng.pep	NLAAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPIALSS	IAYWGLASADRLFLKKY	240
60	orf10nm	NLAAAAFLLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSS	IAYWGLASADRLFLKKY	240
	orf10ng.pep	AGLEQLGVYSMGISFGGAALLFQSI	FSTVWTPYIFRAIEENATPARLSATAESAAALLAS	300
	orf10nm	AGLEQLGVYSMGISFGGAALLFQSI	FSTVWTPYIFRAIEENAPPARLSATAESAAALLAS	300
65	orf10ng.pep	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFYTLTEISGIGLNVVRKTRPIALAT	360
	orf10nm	ALCXTGIFSPLASLLLPENYA	AVRFIVVSCMXPPLFCTLA	ISGIGLNVVRKTRPIALAT 360

-241-

		370	380	390	400	410
5	orf10ng.pep	LGALAANLLLLGL--AVPSGGTRGA	AVACASFWLFFVKTESSCRLWQPLKRLPLYMHT			
	orf10nm	LGALAANLLLLGLDRAVPAR-PXGA	AVACASFWLFFAFKTESSCRLWQPLKRLPLYLHT			
		370	380	390	400	410
10	orf10ng.pep	LFCLASSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKNLHKL FHYLKKQGFPLX			
	orf10nm	LFCLTSSAAYTCFGTPANYPLFAGV	WAAYLAGCILRHRKDLHKL FHYLKKQGFPLX			
		420	430	440	450	470

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

15	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
	51	GGTTTATGCC	GTCATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCcccgCCG
	101	ACGACATCGG	GCGCATCGTG	CTGATGCAGA	CGGCGGCGGG	ACTGACGGTG
	151	TCGGTATTGT	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCTCTG	CCGCCGCTGC
20	251	TGTTTTCGCG	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	GTCCCTGCCG
	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCGCGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTCTCTTA	CTGGTTTTCG
	401	GTATGGAAGG	GCGCGCCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGCCCAAA
	451	CTCGCCATTC	TGCTGCTGTT	GCCGCTGACG	GTCGGGCTGC	TGCACTTTC
	501	GGCGAACACC	TCCGTCCTGA	CCGCCGTTTA	CGCGCTGGCA	AACCTTGCCG
25	551	CCGCCGCTT	TTTGCTGTTT	CAAAACCGAT	GCCGCTGAA	GGCCGTCGGG
	601	CGCGCGCCGT	TTTCGCCCCG	CGTCCTGCAC	CGGGGCTGC	GCTACGGCAT
	651	ACCGCTCGCA	CTGAGCAGCC	TTGCCTATTG	GGGGCTGGCA	TCCGCCGACC
	701	GTTTGTTCTT	GAAAAATAT	GCGGGCCTGG	AACAGCTCGG	CGTTTATTCG
30	751	ATGGGTATTT	CGTTCGGCGG	GCGCGCATTA	TTGCTCCAAA	GCATCTTTTC
	801	AACGGTCTGG	ACACCGTATA	TTTTCCGTGC	AATCGAAGAA	AACGCCACGC
	851	CCGCCGCGCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCTCTC
	901	GCCCTCTGCC	TGACCGGAAT	TTTCTCGCCC	CTCGCTCCC	TCCTGCTGCC
	951	GGAAAACTAC	GCCGCCGTCC	GGTTTACCGT	CGTATCGTGT	ATGCTGccgc
35	1001	cgctGTTTTA	CACGCTGACC	GAAATCAGCG	GCATCGGTTT	GAACGTCGTC
	1051	CGCAAACCGC	GTCCGATCGC	GCTTGCCACC	TTGGGCGCGC	TGGCGGCAAA
	1101	CCTGCTGCTG	CTGGGCTTTC	CCGTACCGTC	CGGCGGCACG	CGCGGCGCGG
	1151	CGGTTGCGCTG	TGCCGCTCA	TTCTGGTTGT	TTTTTGTTTT	CAAGACAGAA
	1201	AGCTCCTGCC	GCCTGTGGCA	GCCGCTCAAA	CGCCTGCCGC	TTTATATGCA
	1251	CACATGTGTC	TGCCTgGCCT	CCTCGGCGGC	CTACACCTGC	TTCGGCACAC
40	1301	CGGCAAACTA	CCCcctggtt	gccggcggtAT	GGGCGGCATA	TCTGGCAGGC
	1351	TGCATCTGCG	GCCACCGGAA	AAATTGTCAC	AAACTGTTTC	ATTATTTGAA
	1401	AAAACAAGGT	TTCCATTAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 380>:

45	1	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV
	51	SVLCLGLDQA	YVREYYAAAD	KDTLFKTLFL	PLLFSAAIA	ALLSRPSLP
	101	SEILFSLDDA	AAGIGLVLFE	LSFLPIRFL	LVLRMEGRAL	AFSSAQLVPK
	151	LAIIIIIIPLT	VGLLHFPANT	SVLTAVYALA	NLAAAFLLF	QNRCLKAVR
	201	RAPFSPAVLH	RGLRYGIPLA	LSSLAYWGLA	SADRLFLKKY	AGLEQLGVYS
50	251	MGISFGGAAL	LLQSIFSTVW	TPYIFRAIEE	NATPARLSAT	AESAAALLAS
	301	ALCLTGIFSP	LASLLLPENY	AAVRETVVSC	MLPPLEYTLT	EISGIGLNVV
	351	RKTRPIALAT	LGALAANLLL	LGLAVPSGGT	RGAAVACAAS	FWLFFVKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAYTC	FGTPANYPLF	AGVWAAYLAG
	451	CILRHRKNLH	KLFHYLKKQG	FPL*		

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

55		10	20	30	40	50	60
	orf10-1.pep	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
	orf10ng-1	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA					
60		10	20	30	40	50	60
	orf10-1.pep	YVREYYATADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
	orf10ng-1	YVREYYAAADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE					
65		70	80	90	100	110	120

-242-

		130	140	150	160	170	180
5	orf10-1.pep	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	LLLLPLTVGLLHFPANTAVLTAVYALA				
	orf10ng-1	LSFLPIRFLLLVLRMEGRALAFSSAQLVPKLA	LLLLPLTVGLLHFPANTSVLTAVYALA				
		130	140	150	160	170	180
10	orf10-1.pep	NLAAAAFLLFQNRCLKAVRHAPFSPAVLHRGLRYGIPIALSS	IAYWGLASADRLFLKKY				
	orf10ng-1	NLAAAAFLLFQNRCLKAVRRAPFSPAVLHRGLRYGIPLALSS	LAYWGLASADRLFLKKY				
		190	200	210	220	230	240
15	orf10-1.pep	AGLEQLGVYSMGISFGGAALLQSFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS					
	orf10ng-1	AGLEQLGVYSMGISFGGAALLQSFSTVWTPYIFRAIEENATPARLSATAESAAALLAS					
		250	260	270	280	290	300
20	orf10-1.pep	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFCTLA	EISGIGLN	VVRKTRPIALAT		
	orf10ng-1	ALCLTGIFSPLASLLLPENYA	AVRFIVVSCMLPPLFYTLTEISGIGLN	VVRKTRPIALAT			
		310	320	330	340	350	360
25	orf10-1.pep	LGALAANLLLLGLAVPSGGARGA	AVACAASFWLFFFAKTESSCRLWQPLKRLPLYLHTLF				
	orf10ng-1	LGALAANLLLLGLAVPSGGTRGA	AVACAASFWLFFFAKTESSCRLWQPLKRLPLYMHTLF				
		370	380	390	400	410	420
30	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVW	AAYLAGCILRHRKDLHKLFHYLKKQGFPLX				
	orf10ng-1	CLASSAAYTCFGTPANYPLFAGVW	AAYLAGCILRHRKNLHKLFHYLKKQGFPLX				
		430	440	450	460	470	
35	orf10-1.pep	CLTSSAAYTCFGTPANYPLFAGVW	AAYLAGCILRHRKDLHKLFHYLKKQGFPLX				
	orf10ng-1	CLASSAAYTCFGTPANYPLFAGVW	AAYLAGCILRHRKNLHKLFHYLKKQGFPLX				
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several
 40 transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6
 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and
 their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCCTGAAAC CGCATAACCA GCTTAAGGAA GACATCCAAC CTGATCCGGC
	51 CGATCAAAAC GCCTTGTC CG AACC GGATGC TGCGACAGAG GCAGAGCAGT
	101 CGGATGCGGA AAATGCTGCC GACAAGCAGC CCGTTGCCGA TAAAGCCGAC
	151 GAGGTTGAAG AAAAGGCGGG CGAGCCGGAA CGGGAAGAGC CGGACGGACA
	201 GGCAGTGCGT AAGAAAGCGC TGACGGAAGA GCGTGAACAA ACCGTCAGGG
50	251 AAAAAGCGCA GAAGAAAGAT GCCGAAACGG TTAAATACA AGCGGTAAAA
	301 CCGTCTAAAG AAACAGAGAA AAAAGCTTCA AAAGAAGAGA AAAAGGCGGC
	351 GAAGGAAAAA GTTGCACCCA AACCAACCCC GGAACAAATC CTCAACAGCG
	401 GCAgCATCGA AAaMGCGCG AgTGCCGCGG CCAAAGAAGT GCAGAAAAATG
	451 AA.AACGTCC GACAAGGCGG AAGC.AACGC ATTATCTGCA AATGGGCGCG
55	501 TATGCCGACC GTCAGAGCGC GGAAGGGCAG CGTGCCAAAC TGGCAATCTT
	551 GGGCATATCT TCCAAGGTGG TCGGTATCA GGCGGGACAT AAAACGCTTT
	601 ACCGGGTGCA AAGCGGCAAT ATGCTGCCG ATGCGGTGA

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKPHNLKE DIQDPADQN ALSEPDAATE AEQSDAENAA DKQPVADKAD
	51 EVEEKAGEPE REEPDQAVR KKALTEEREQ TVREKAQKKD AETVKIQAVK

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N.*

meningitidis:

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
51 CTTGCGTTTG ATACTGGCGA CGGTCAATTAT TGCCGGTATT TTGTTTTATC

-244-

101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGTTCC GTCGAAGCAG
 151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
 5 251 AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
 301 GCCGACAAAAG CCGACGAGGT TGAGGAAAAG GCGGACGAGC CGGAGCGGGA
 351 AAAGTCGGAC GGACAGGCAG TGGCAAGAA AGCACTGACG GAAGAGCGTG
 401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACGGTTAAA
 451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAAAG
 10 501 AGAGAAAAAG GCGGAGAAGG AAAAAGTTGC ACCCAAACCG ACCCCGGAAC
 551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCAAA
 601 GAAGTCGAGA AAATGAAAAC GCCCGACAAG GCGGAAGCAA CGCATTATCT
 651 GCAAAATGGC GCGTATGCCG ACCGCCGAG CGCGGAAGG CAGCGTGCCA
 701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA
 751 CATAAAACGC TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
 15 801 GAAAAAATG CAGGACGAGT TGAAAAACA TGAAGTCGCC AGCCTGATCC
 851 GTTCTATCGA AAGCAAATAA

This encodes a protein having amino acid sequence <SEQ ID 386>:

1 MFMNKFQSG KGLSGFFFL ILATVIIAGI LFYLNQSGQN AFKIPVPSKQ
 51 PAETEILKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV
 101 ADKADEVEEK ADEPEREKSD QGAVRKALTE EEREQTVGEK AQKKDAETVK
 151 KQAVKPSKET EKKASKEEKK AEKEKVAPKP TPEQILNSGS IEKARSAAK
 201 EVQKMKTPDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGQYQAG
 251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25	orf65a.pep	10	20	30	40	50	60
	orf65-1	10	20	30	40	50	60
30	orf65a.pep	70	80	90	100	110	120
	orf65-1	70	80	90	100	110	120
35	orf65a.pep	130	140	150	160	170	180
	orf65-1	130	140	150	160	170	180
40	orf65a.pep	190	200	210	220	230	240
	orf65-1	190	200	210	220	230	240
45	orf65a.pep	250	260	270	280	290	
	orf65-1	250	260	270	280	290	

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60	ORF65ng	30	40	50	60	70	80
	ORF65	10	20	30			

-245-

		90	100	110	120	130	140
ORF65ng	AEQSDAEKAADKQPVADKADEVEEKAGEPEREEPDQAVRKKALTEEREQTVREKAQKKD						
	: : : : : : :						
ORF65	AEQSDAENAADKQPVADKADEVEEKAGEPEREEPDQAVRKKALTEEREQTVREKAQKKD						
5		40	50	60	70	80	90
		150	160	170	180	190	200
ORF65ng	AETVKKKAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM						
	: : : : : : :						
10	ORF65	AETVKIQAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEKARSAAAKEVQKM					
		100	110	120	130	140	150
		210	220	230	240	250	260
ORF65ng	KNFGQGGSQRIICKWARMNPNGARKGSPVNWQSWAYLPKWSAIRRDIKRETACKAAICPP						
15	ORF65	XNVRGGXSRIICKWARMPTVRARKGSPVNWQSWAYLPKWSAIRRDIKRETGCKAAICLP					
		160	170	180	190	200	210
ORF65ng	MR						
20	ORF65	MR					

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

1	MEMNKFSQSG	KGLSGFFFL	ILATVILIAGI	LLYLNQGGQN	AFKIPAPSKQ
51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
151	KKAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
201	EVQKMKNFQ	GGSQRIICKW	ARMPNPGARK	GSPVNWQSWA	YLPKWSAIRR
251	DIKRFTACKA	AICPPMR*			

After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

1	ATGTTTATGA	ACAAATTTTC	CCAATCCGGA	AAAGGTCTGT	CCGGTTTCTT
51	CTTCGGTTTG	ATACTGGCAA	CGGTCATTAT	TGCCGGTATT	TTGCTTTATC
101	TGAACCAGGG	CGGTCAAAAT	GCGTTCAAAA	TCCCGGTCC	GTCGAAGCAG
151	CCTGCAGAAA	CGGAAATCCT	GAAACTGAAA	AACCAGCCTA	AGGAAGACAT
201	CCAACCTGAA	CCGGCCGATC	AAAACGCCTT	GTCCGAACCG	GATGTTGCGA
251	AAGAGGCAGA	GCAGTCGGAT	GCGGAAAAAG	CTGCCGACAA	GCAGCCCGTT
301	GCCGACAAag	ccgacgAGGT	TGAAGAAAag	GcGGgcgAgc	cggaACGGga
351	aGAGCCGGAC	ggACAGGCAG	TGCGCAAGAA	AGCACTGAcg	gAAGAgcGTG
401	AACAAACcgt	cagggAAAAA	GCGCagaaga	AAGATGCCGA	AACGgTTAAA
451	AAacaaGCgg	tAaaaccgtc	tAAAGAAACa	gagaaaaaag	cTtcaaaaga
501	agagaaaaag	gcggcgaaaag	aaaAAGttgc	acccaaaccg	accccggaac
551	aatcctctcaa	cagccgCagc	atcgaaaaag	cgcgtagtgc	cgctgccaaa
601	gaAgTgcagA	AAatgaaaaa	ctTtgggcaa	ggcgGaaagc	aacgcattaT
651	CTGcaaatgg	gcgcgtatgc	cgaccgtccg	gagcgcggaA	gggcagcgtg
701	ccaaACtggc	aAtcttgGgc	atatctTccg	aagtgtcgcG	CTATCAGGCG
751	GGACATAAAA	CGCTTTACCG	CGTGCAaagc	GGCAatatgt	ccgccgatgc
801	gGTGAAAAAA	ATGCAGGACG	AGTTGAAAAA	GCATGGGGtt	gcCAGCCTGA
851	TCCGTGcgAT	TGAAGCAAAA	TAA		

This encodes the following amino acid sequence <SEQ ID 390>:

1	MEMNKFSQSG	KGLSGFFFL	ILATVILIAGI	LLYLNQGGQN	AFKIPAPSKQ
51	PAETEILKLK	NQPKEDIQPE	PADQNALSEP	DVAKEAEQSD	AEKAADKQPV
101	ADKADEVEEK	AGEPEREEDP	GQAVRKKALT	EEREQTVREK	AQKKDAETVK
151	KQAVKPSKET	EKKASKEEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
201	EVQKMKNFQ	GGSQRIICKW	ARMPTVRSAE	GQRAKLAILG	ISSEVVGYYA
251	GKHTLYRVQS	GNMSADAVKK	MQDELKKHGV	ASLIRAIEGK	*

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

		10	20	30	40	50	60
orf65-1.pep	MEMNKFSQSGKGLSGFFFLILATVILIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK						
	: : : : : : :						
60	orf65ng-1	MEMNKFSQSGKGLSGFFFLILATVILIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK					
		10	20	30	40	50	60

-246-

		70	80	90	100	110	120
	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD					
5	orf65ng-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEPD					
		70	80	90	100	110	120
	orf65-1.pep	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETTEKKASKEEKKAAKEKVAPKP					
10	orf65ng-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETTEKKASKEEKKAAKEKVAPKP					
		130	140	150	160	170	180
	orf65-1.pep	TPEQILNSGSIEKARSAAAKEVQKMKTSKAEATHYL-QMGAYADRQSAEGQRAKLAILG					
15	orf65ng-1	TPEQILNSRSIEKARSAAAKEVQKMKNFQGGSQRIICKWARMPTVRSAGQRAKLAILG					
		190	200	210	220	230	239
	orf65-1.pep	ISSKVVGYYQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
20	orf65ng-1	ISSKVVGYYQAGHKTLRYVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		240	250	260	270	280	290
	orf65-1.pep	ISSKVVGYYQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65ng-1	ISSKVVGYYQAGHKTLRYVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		250	260	270	280	290	

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 391>:

30	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCTACTCG	GTkTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGs.s
	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
35	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCCGATTAAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTtaggCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAaATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTtACCCATA	AAATCCATAC
40	401	CGGCCTGCCT	tGCGgTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GCTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AgCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCTT	TTGCACTGGG	TACGCTGCCC	AATCTTtTAG
	551	CAATCGGCAT	TTTtTCCCTG	CAACTGAaWA	AAATCATGCA	AAACCGATAT
	601	ATCCGCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGCTCTG	TGGCTGTAA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

	1	MNHDITFTL	FLLGXFGGTH	CIGMCGGLSS	AFXXQLPPhi	NRFWLILLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLEFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPi	KSIPACLAVG	ILWGWLPCGL
	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLp	NLLAIGIFSL	QLXKIMQNRy
50	201	IRLCTGLSVS	LWLWKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
55	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCCGATTAAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTtaggCTT	ATACTTGAGC
	301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAaATCGGCA	AACCGATATG

-247-

5
351 GCGGAACCTG AACCCGATAC TCAACCGCT GTTACCCATA AAATCCATAC
401 CCGCCTGCCT TCGGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTAG
551 CAATCGGCAT TTTTCCCTG CAACTGAAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAACT
651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

10
1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NrfWLILLN
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNilyTAANL LLLFLGLYLS
101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACIavg ILWGWLPCGL
151 VYSASLYALG SGSAAATGGly MLAFALGTLP NLLAIGIFSL QLKIMQNRy
201 IRLCTGLSVS LWALWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N.*

meningitidis:

20
orff103.pep 10 20 30 40 50 60
MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPhINrfWLILLNLTGRVSSYTAI
orff103a 10 20 30 40 50 60
MNxDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPhINrxWLILLNLTGRVSSYTAI
25
orff103.pep 70 80 90 100 110 120
GLILGLIGQVGVSLDQTRVLQNilyTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
orff103a 70 80 90 100 110 120
GLILGLIGQVGVSLDQTRVXQNilyTAANLLLFLGLYLSGISSLAAKIEKIGKPIWRNL
30
orff103.pep 130 140 150 160 170 180
NPILNRLLPIKSIPACIavgILWGWLPCGLVYSASLYALGSGSAATGGlyMLAFALGTLP
orff103a 130 140 150 160 170 180
NPILNRLLPIKSIPACIavgILWGWLPCGLVYSASLYALGSGSAATGGlyMLAFALGTLP
35
orff103.pep 190 200 210 220
NLLAIGIFSLQLXKIMQNRyIRLCTGLSVSLWALWKLAVLWLX
orff103a 190 200 210 220
NLXAIGIFSLQLXKIMQNRyIRLCTGLSVSLWALWKLAVLWLX
40

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

45
1 ATGAACCANG ACATCACTTT CCTCACCTG TTCCTACTCG GTTCTTCGG
51 CGGAACGCAC TGCATCGGTA TGTGCGGCGG ATTAAGCAGC GCGTTTGCGC
101 TCCAACCTCC CCCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC
151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT
201 CGGACAGGTC GCGCTTTCAC TCGACCAAAC CCGCGTCNTG CAGAATATT
251 TATACACGGC CGCCAACTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
301 GGTATTCTT CTTGGCGGC AAAAATCGAG AAAATCGGCA AACCGATATG
351 GCGGAACCTG AACCCGATAC TCAACCGCT GTTACCCATA AAATCCATAC
401 CCGCCTGCCT TCGGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA
451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTNGG
551 CAATCGGCAT TTTTCCCTG CAACTGNAAA AAATCATGCA AAACCGATAT
601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAACT
55
651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNxDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPhi NRxWLILLN
51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNilyTAANL LLLFLGLYLS
101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACIavg ILWGWLPCGL

151 VYSASLYALG SGSAATGGLY MLAFALGTLN NLXAIGIFSL QLXKIMQNRY
 201 IRLCTGLSVS LWALWKLAVL WL*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

		10	20	30	40	50	60
5	orf103a.pep	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLIIILLNTGRVSSYTAI					
	orf103-1	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLIIILLNTGRVSSYTAI					
		10	20	30	40	50	60
10	orf103a.pep	GLILGLIGQVGVSLDQTRVQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103-1	GLILGLIGQVGVSLDQTRVQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
15	orf103a.pep	NPILNRLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLN					
	orf103-1	NPILNRLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLN					
		130	140	150	160	170	180
20	orf103a.pep	NLXAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		
25	orf103a.pep	NLXAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
	orf103-1	NLLAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWLX					
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFWLIIILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLIIILLNTGRVSSYTAI	60
35	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
40	orf103.pep	NPILNRLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSAATGGLYMLAFALGTLN	180
	orf103ng	NPILNRLPIKSIPACLA VGILWGWLPCGLVYSASLYALGSGSATTGGLYMLAFALGTLN	180
	orf103.pep	NLLAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWL 222	
45	orf103ng	NLLAIGIFSLQLXKIMQNRYIRLCTGLSVSLWALWKLAVLWL 222	

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

	1	ATGAACCACG	ACATCACTTT	CCTCACCTCG	TTCCTGCTCG	GTTTCTTCGG
	51	CGGAACACAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCCTTTGCGC
50	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
	151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
	201	CGGACAACCT	GGCATTTCAC	TCGACCAAAc	ccgcgTCCTG	CAAAATATTT
	251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTtaggCTT	ATACTTGAGC
	301	gGTATTCTTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GGCACAACCT	AACCCGATAC	TCAACCGGCT	GCTGCCCATG	AAATCCATAC
55	401	CCGCCTGCCT	TGCTGTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
	451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
	501	CGGACTGTAT	ATGCTTGCTT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
	551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
60	651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

-249-

1 MNHDITFTL FLLGFFGGTH CIGMCGGLSS AFALQLPPI NRFWLILLN
 51 TGRISSTAI GLMLGLIGQL GISLDQTRVL QNILYASNL LLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NPILNRLLPI KSIPACLAVG ILWGWLP
 151 VYSASYALG SGSATTGGLY MLAFALGTL NLLAIGIFSL QLKKIMQNR
 201 IRLCTGLSVS LWALWKLAVL WL*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
10	orf103-1.pep	MNHDITFTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NRFWLILLN	TGRVSSYTAI
	orf103ng	MNHDITFTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPI	NRFWLILLN	TGRISSTAI
		10	20	30	40	50	60
15	orf103-1.pep	GLILGLIGQ	VGVSLDQTRVL	QNILYTAAN	LLFLGLYLS	GISSLAAKIE	KIGKPIWRNL
	orf103ng	GLMLGLIGQ	LGISLDQTRVL	QNILYASNL	LLFLGLYLS	GISSLAAKIE	KIGKPIWRNL
		70	80	90	100	110	120
20	orf103-1.pep	NPILNRLLPI	KSIPACLAVG	ILWGWLP	PCGLVYSAS	LYALGSGSA	ATGGGLYMLAFALGTL
	orf103ng	NPILNRLLPI	KSIPACLAVG	ILWGWLP	PCGLVYSAS	LYALGSGSA	ATGGGLYMLAFALGTL
		130	140	150	160	170	180
25	orf103-1.pep	NLLAIGIFSL	QLKKIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
	orf103ng	NLLAIGIFSL	QLKKIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
		190	200	210	220		

30 Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

	1	ATGGA	AAACC	AAAGGCCGCT	CCTAGGCTTT	CGCTTGGCAC	TTTGGCGGC
	51	GATGACGTGG	GGAACGCTGC	CGAT.TCCGT	GCGGCAGGTA	TTGAAGTTTG	
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA	
40	151	TTGTTTGT	TTTGCTGGCACT	GGGCGGCGG	CTGCGGAGC	GGCGaGGATT	
	201	TTTCTTGGTG	CTCATTTCAGG	CTGCTGCTGC	TCGGCGTGGC	GGGCATTTCG	
	251	GCAAACTTTG	TGCTGATTGC	CCAAGGCTG	CATTATATTT	CGCCGACCAC	
	301	GACGCAGGTT	TTGTGGCAGA	TTTCGCCGTT	TACGATGATT	GTWGTGGTG	
	351	TGTTTGGTGT	TAAAGACCGG	ATGACTGCCG	CTCAGAAAAT	CGGCTTGGTT	
45	401	TTGCTGCTTG	CCGGTTTGCT	TATGTATTTT	AACGATAAAT	TCGGCGAGTT	
	451	GTCGGGTTTG	GGCGCGTATG	C.AAGGGCGT	GTTGCTGTGT	GCGGCAGGCA	
	501	GTATGGCATG	GGTGTGTAAT	GCCGTGGCGC	AAAAGCTGCT	GTGCGCGCAA	
	551	TTCCGGCCGC	AACAGATTCT	GCTGTGATT	TATGCGGCAA	GTGCGCGCGT	
	601	GTTCTGCGC	TTTGCCGAAC	CGGCACACAT	CGGAAGTATG	GACGGTACGT	
	651	TGGCGTGGGT	ATGTATTGCG	TATTGCTGCT	TGAATACGTT	AATCGGTTAC	
50	701	GGCTCGTTCG	GCGAGGCGTT	GAAACATTGG	GAGGCTTCCA	AAGTCAGCGC	
	751	GGTAACAACC	TTGCTCCCGG	TGTTTACCGT	AATAAATACT	TTGCTCGGGC	
	801	ATTATGTGAT	GCCTGAAACT	TTTGCCGCGC	CGGA..		

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

	1	MENQR	PLLGF	RLALLAAMTW	GTLPXSVRQV	LKFVDAPTLV	WVRFTVAAAV
55	51	LFVLLALGGR	LPKRRDFSWC	SFRLLLLGVA	GISANFVLIA	OGLHYISPTT	
	101	TQVLWQISPF	TMIVVGVLVF	KDRMTAAQKI	GLVLLLAGLL	MYFNDKFGEL	
	151	SGLGAYXKGV	LLCAAGSMAW	VCONAVAQKLL	SAQFGPQQIL	LLIYAASAAV	
	201	FLPFAEPAHI	GSMGDTLAWV	CIAYCCLNTL	IGYGSFGEAL	KHWEASKVSA	

	1	ATGGA AAAACC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
5	51	GATGACGTGG	GGAACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGCC	GGCGCGCGTA
	151	TTGTTTGTTT	TGCTGGCACT	GGCGGGCGGG	TCGCCGAAGC	GGCGGGATTT
	201	TTCTTGGTGC	TCATTAGGC	TGCTGCTGCT	CGGCGTGGCG	GGCATTTCGG
	251	CAAACTTTGT	TGCTGATTGCC	CAAGGGCTGC	ATTATATTTT	GCCGACCACG
10	301	ACGCAGGTTT	TGTGGCAGAT	TTCCGCCGTT	ACGATGATTG	TTGTGCGGTG
	351	GTTGGTGTTT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGGTTT
	401	TGCTGCTGCT	CGGTTTGCTT	ATGTTTTTTA	ACGATAAATT	CGGCGAGTTG
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TGTGCTGTGT	CGCGAGGCAG
	501	TATGGCATGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
15	551	TCGGGCGCGA	ACAGATTCTG	CTGTTGATTT	ATGCGGCAAG	TGCCGCGGTG
	601	TTCTCTGCCG	TGCGCGAACC	GGCACACATC	GGAAGTTTGG	ACGGTACGTT
	651	GGCGTGGGTT	TGTTTGTGCT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
	701	GCTCGTTCCG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCCGT	GTTTACCGTA	ATAwTwwCTT	TGCTCGGGCA
	801	TTATGTGATG	CTGAAACTT	TTGCCGCGCC	GGG...	

1	MENQRPLLG	ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
51	LFVLVLALGGR	LPKKRDFSCV	SRFLLLLGVA	GISANFVLIA	QGLHYISPFT
101	TOVLVQISPE	TMVVGVLVF	KDRMTAAQKI	GLVLLLAGLL	MFENDKSGTT
151	SGLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQOIL	LLIYAASAAV
201	FLPFPEPAHI	GSLDGLTAW	CFYACCLNTL	IGYGSFGEAL	KHWEASKVSA
251	VTTLPVPVTF	ISXLLGHYVM	PETAFA...		

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

30	orf104	4	QRPLLGLFRLLALLAAMTWGTLPKXSVRQVLKEVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62
			Q+PLLGF AL+ AM WG+LP +++QVL ++A T+VW P
	HI0878	3	QQPLLGFTFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62
35	orf104	63	--KRRDfSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTQVLWQISPFTMIVGVLVF 120
			K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
	HI0878	63	LMKVRQYAW----IMLIGVIGLTSNFLLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLIF 118
40	orf104	121	KDRMTAAQKIXXXXXXXXXXXMYFNDKFGEISGLGAYXKGVLCAAGSMAWVCNAVAQKLL 180
			K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
	HI0878	119	KEKLGLHQKIGLFLLLIGLGLFFNDRFDAFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178
45	orf104	181	SAQFGPQQILLLIYAASAAVFLPFAEPAHIGSMOGLTAWVCIAYCCLNTLIGYGSFGEAL 240
			+F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL
	HI0878	179	LRKFNSQQILLMMYLGCAIAFMFMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237
	orf104	241	KHWEASKVSAVTLLPVFTVINTLLGHYVMPETFAAP 277
			NW+ SKVS V TL+P+FT++ + + HY P FAAP
	HI0878	238	LRWDVSKVSVVITLVPLETILFESHIAHYFSPADFAAP 274

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

		10	20	30	40	50	60
55	orf104.pep	MENQRPLLGFRLLAAMTGWGTLEPXSVRQVLKFVDAPTLVWVRFVAAAVLFVLLALGGR					
	orf104a	MENQRPLLGFRLLAAMTGWGTLPVAVRQVLKFVDAPTLVWVRFVAAAVLFVLLALGGR					
		10	20	30	40	50	60
		70	80	90	100	110	120

-251-

```

5  orf104.pep  LPKRRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   orf104a    LPKWRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
                                     70      80      90      100     110     120

10  orf104.pep  KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
   orf104a    KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
                                     130     140     150     160     170     180

15  orf104.pep  SAQFGPQQILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIAYCCLNTLIGYGSFGEAL
   orf104a    SAQFGPQQILLIYAASAAVFLPFAELAHIGSLDGTLAWVCFAYCCLNTLIGYGSFGEAL
                                     190     200     210     220     230     240

20  orf104.pep  KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP
   orf104a    KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLYAGALVVVGGAVTAAVG
                                     250     260     270     280     290     300

```

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

```

25  1  ATGGAACACC AAAGGCCGCT CCTAGGCTTC GCGTTGGCAC TTTTGGCGGC
   51  GATGACGTGG GGAACGCTGC CGATTGCCGT GCGGCAGGTA TTGAAGTTTG
  101  TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GGCGGCGGTA
  151  TTGTTTGTTC TGCTGGCATT GGGCGGGCGG CTGCCGAAGT GGCGGGATTT
  201  TTCTTGGTGC TCATTAGGC TGCTGTGCTG CGGCGTGGCG GGCATTTCCGG
  251  CAAACTTTGT GCTGATTGCC CAAGGGCTGC ATTATATTTT GCCGACCACG
  301  ACGCAGGTTT TGTGGCAGAT TTCGCCGTTT ACGATGATTG TTGTCGGTGT
  351  GTTGGTGTTC AAAGACCGGA TGAATGCCGC TCAGAAATC GGCTTGGTTT
  401  TGCTGCTTGC CGGTTTGCTT ATGTTTTTTA ACGATAAATT CGGCGAGTTG
  451  TCGGGTTTGG GCGCGTATGC GAAGGGCGTG TTGCTGTGTG CGGCAGGCAG
  501  TATGGCATGG GTGTGTTATG CCGTGGCGCA AAAGCTGCTG TCGGCGCAAT
  551  TCGGGCCGCA ACAGATCTCT CTGTTGATTG ATGCGGCAAG TGCCGCCGTG
  601  TTCCTGCCGT TTGCCGAAC GGCACACATC GGAAGTTTGG ACGGTACGTT
  651  GGCCTGGGTT TGTGTTGCGT ATTGCTGCTT GAATACGTTA ATCGGTTACG
  701  GCTCGTTTCG CGAGGCGTTG AAACATTGGG AGGCTTCCAA AGTCAGCGCG
  751  GTAACAACCT TGCTCCCGCT GTTTACCGTA ATATTTCTT TGCTCGGGCA
  801  TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTGTTGGGT
  851  ATGCCGCGCG ACTGGTCGTG GTCGGGGGTG CGGTTACGGC GGCGGTGGGG
  901  GACAGGCTGT TCAAACGCCG CTAG

```

This encodes a protein having amino acid sequence <SEQ ID 404>:

```

45  1  MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
   51  LFLVLLALGGR LPKWRDFSWS SFRLLLLGVA GISANFVLI A QGLHYISPTT
  101  TVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MFFNDKFGEL
  151  SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
  201  FLPPFAELAH GSLDGTLAWV CFAYCCLNTL IGYGSFGEAL KHWEASKVSA
  251  VTTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLYAGALVV VGGAVTAAVG
  301  DRLFKRR*

```

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

```

55  orf104a.pep  MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR
   orf104-1     MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR
                                     10      20      30      40      50      60

60  orf104a.pep  LPKWRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
   orf104-1     LPKRRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF
                                     70      80      90      100     110     120

65  orf104a.pep  KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL

```

-252-

	orfl04-1	 KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	130	140	150	160	170	180
5	orfl04a.pep	SAQFGPQQIILLIYAASA AVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220	230	240
	orfl04-1	 SAQFGPQQIILLIYAASA AVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL	190	200	210	220	230	240
10	orfl04a.pep	 KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYAGALVVVGAVTA AVG	250	260	270	280	290	300
15	orfl04-1	 KHWEASKVSAVTTLLPVFTVIXLLGHYVMPETFAAP	250	260	270			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orfl04.pep	MENQRPLLGFRLALLAAMTWGTL PKXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orfl04ng	: MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
25	orfl04.pep	LPKRRDFSWCSFRLLLLVAGISANFVLIAQGLHYISPTTQVLWQISPFMTIVVGVLVF	120
	orfl04ng	: LPKRRDFSWHSFRLLLLVGTGISANFVLIAQGLHYISPTTQVLWQISPFMTIVVGVLVF	120
30	orfl04.pep	KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180
	orfl04ng	: KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orfl04.pep	SAQFGPQQIILLIYAASA AVFLPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL	240
	orfl04ng	: : SAQFGPQQIILLIYAASA AVFLXAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL	240
	orfl04.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP	277
	orfl04ng	: : KHWEASKVSAVTTLLPVFTVIFSL LGHYVMPDTFAAPDMNGLGYVAGALVVVGAVTA AVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

1	MENQRPLLGF ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
51	LFVLLALGGR LPKRRDFSWH	SFRLLLLVGT	GISANFVLIA	QGLHYISPTT
101	TQVLWQISPF TMIVVGVLVF	KDRMTAAQKI	GLVLLLVGLL	MFNDKFGEL
151	SGLGAYAKGV LLCAAGSMAW	VCIYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
201	FLLXAEPAHI GSLDGT LAWV	CFVYCCLNTL	IGYGSFGEAL	KHWEASKVSA
251	VTTLLPVFTV IFSLLGHYVM	PDTFAAPDMN	GLGYVAGALV	VVGAVTA AVG
301	DRPFKRR*			

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

1	ATGGA AACC	AAAGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTTGGCGGC
50	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA
	101	TCCATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC
	151	TTGTTTGT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGC
	201	TTCTTGGCAT	TCATTTCAGGC	TGCTGCTGCT	CGGCGTGACG
55	251	CAAACTTTGT	GCTGATTGCC	CAAGGCTGCT	ATTATATTTC
	301	ACGCAGGTTT	TGTGCGAGAT	TTCGCCGTTT	ACGATGATTG
	351	GTTGGTGT	AAAGACCGGA	tgaCTGCCGC	GCAGAAAATC
	401	TGCTGCTtgT	CGGTtgCTT	ATGTTTtTta	ACGACAAAT
	451	TCGGGTTTGG	GCGCGTATGC	GAAGGCGGTG	TTGCTGTGTG
	501	TATGCGCTGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG
60	551	TCGGGCGGCA	ACAGATTCTG	CTGTTGATT	ATGCGGcaag
	601	TTCTtgccgT	TTGccgaaCC	GGCACACATC	GGAAGTTTg
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA

701 GCTCGTTCGG CGAGGCGTTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACAACCT TGCTCCCCGT GTTTACCGTA ATATTTCTT TGCTCGGGCA
 801 TTATGTGATG CCGTACTT TGGCCGCCG GGATATGAAC GGTGTGGGT
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGTG CGGTACGGC GCGGTGGG
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFLVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPPFAEPAHI GSLDGT LAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGVGALVV VGGAVTAAVG
 301 DRPFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFLVLLALGGR						
	orf104ng-1	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFDAPTLVWVRFTVAAAVLFLVLLALGGR						
20	orf104-1.pep	70	80	90	100	110	120
	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF						
	orf104ng-1	70	80	90	100	110	120
	LPKRRDFSWHSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF						
25	orf104-1.pep	130	140	150	160	170	180
	KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL						
	orf104ng-1	130	140	150	160	170	180
	KDRMTAAQKIGLVLLLVGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL						
30	orf104-1.pep	190	200	210	220	230	240
	SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL						
	orf104ng-1	190	200	210	220	230	240
	SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL						
35	orf104-1.pep	250	260	270			
	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP						
	orf104ng-1	250	260	270	280	290	300
	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGVGALVVVGGAVTAAVG						

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 88
 Q+P M WG+LPIA++QVL ++A T+VW P

Sbjct: 3 QQPLLGFTEFALITAMAWGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

Query: 89 --KRRDFSWHSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFMTIVVGVLVF 146
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+ F

Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNELLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI 118

Query: 147 KDRMTAAQKIXXXXXXXXXXMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
 K+++ QKI +FFND+ F +GL Y+ GV+L G++ WV Y +AQKL+

Sbjct: 119 KEKLGLHQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGALIWVAYGMAQKLM 178

Query: 207 SAQFGPQQIILLIYAASAAVFLPPFAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL 266
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCLNTLIGYGS+ EAL

Sbjct: 179 LRKFNSQQILLMYLGCAIAFMPMADFSQVQELT-PLALICFIYCCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+ET++FS + HY P FAAP++N
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1  ATGGTAGCTC  GTCGGGCTCA  TAACCCGAAG  GTCGTAGGTT  CGAATCCTGT
      51  .CCCGCAACC  TAATTTCAAA  CCCCTCGGTT  CAATGCCGAG  GG.GTTTTGT
     101  T.TTGCCTGT  TTCCTGTTTC  CTGTTTCCTG  CCGCCTCCGT  TTTTGTCCGG
     151  ATTTTCCTTC  CGGCCGCAAT  ATCGGAACGG  CAGACCGCCG  TCTGTTTGCG
     201  GTTGCAAAAT  CAGGCAGTTT  GGCTACAATC  TTCCGCATGT  TCTTCAAGAA
     251  AGCCAACCAT  GCCGACCGTC  CGTTTTACCG  AATCCGTCAG  CAAACAAGAC
     301  CTTGATGCTC  TGTTCGAGTG  GGCAAAAGCA  AGTTACGGTG  CAGAAAAGTTG
     351  CTGGAAAACG  CTGTATCTGA  ACGGTCysCC  TTTGGGCAAC  CTGTGCCCGG
     401  AATGGGTGGA  ACGCGTsmmA  AAAGACTGGG  AGGCAGGCTG  CyCGGAGTCT
     451  TCAGACGGCA  TTTTCTGAA  TgCGGACGGc  TGgCctGATA  TGGgCGGAcg
     501  cTTACAGCAC  CTCGCCCTCG  GTTGGCACTG  TCGGGGGCTG  TTGGACGgsT
     551  GCGCAACGA  GTGTTTCGAC  CTGACCGACG  GCGGCGGCAA  CCCCTTGTTT
     601  ACGCTCGaAc  GCGCCGyTTT  mCGTCCTkTC  GGACTGCTCA  GCCGCGCCGT
     651  CCATCTCAAC  GGTCTGACCG  AATCGGACGG  CCATGGCAT  TTCTGGATAG
     701  CGAGGCGCAG  TCCGCACAAA  GCAGTCGATC  CCAACAAACT  CGACAATACT
     751  rCCGCCGGCG  GTGTTTCCGG  CGGCGAAATG  CCGTCTGAAG  CCGTGTGTCG
     801  CGAAAGCAGC  GAAGAAGCCG  GTTGGATAAA  AACGCTGcTT  CCGCTCATCC
     851  GCCCGGTATC  GCAGCTGCAC  AGCCTGCGCT  CCGTCAGCCG  GGGTGTACAC
     901  AATGAAATCC  TGTATGTATT  CGATGCCGTC  CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1  MVARRAHNPK  VVGSNXPXAT  XFQTPRFNAE  XVLXLPVSCF  LFPAASVFCR
      51  IFLPAAISER  QTAVCLRLQI  QAVWLQSSAL  SSRKPTMPTV  RFTESVSKQD
     101  LDALFEWAKA  SYGAESCWKT  LYLNGLPLGN  LSPWVERVX  KDWEAGCKES
     151  SDGIFLNADG  WPDMMGRLQH  LALGWHCAGL  LDGWRNECFD  LTDGGGNPLF
     201  TLERAXRFX  GLLSRAVHLN  GLTESDGRWH  FWIGRRSPHK  AVDPNKLDNT
     251  XAGGVSGGEM  PSEAVCRESS  EEAGLDKTL  PLIRPVSQLH  SLRSVSRGVH
     301  NEILYVFDAV  LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1  ATGCCGACCG  TCCGTTTTAC  CGAATCCGTC  AGCAAACAAG  ACCTTGATGC
      51  TCTGTTTCGAG  TGGGCAAAAG  CAAGTTACGG  TGCAGAAAGT  TGCTGGAAAA
     101  CGCTGTATCT  GAACGGTCTG  CCTTTGGGCA  ACCTGTCGCC  GGAATGGGTG
     151  GACGCGTCA  AAAAAGACTG  GGAGGCAGGC  TGCTCGGAGT  CTTCAGACGG
     201  CATTTTTCTG  AATGCGGACG  GCTGGCCTGA  TATGGGCGGA  CGCTTACAGC
     251  ACCTCGCCCT  CGGTTGGCAC  TGTGCGGGGC  TGTGGACGG  CTGGCGCAAC
     301  GAGTGTTCG  ACCTGACCGA  CGGCGGCGGC  AACCCTTGT  TCACGCTCGA
     351  ACGCGCCGCT  TTCCGTCTCT  TCGGACTGCT  CAGCCGCGCC  GTCCATCTCA
     401  ACGGTCTGAC  CGAATCGGAC  GGCCGATGGC  ATTTCTGGAT  AGGCAGGCGC
     451  AGTCCGCACA  AAGCAGTCGA  TCCCAACAAA  CTCGACAATA  CTGCCGCCGG
     501  CGGTGTTTCC  GGGCGGCAAA  TGCCGTCTGA  AGCCGTGTGT  CGCGAAAGCA
     551  GCGAAGAAGC  CGGTTTGAT  AAAACGCTGC  TTCCGCTCAT  CCGCCCGGTA
     601  TCGCAGCTGC  ACAGCCTGCG  CTCCGTGACG  CGGGGTGTAC  ACAATGAAAT
     651  CCTGTATGTA  TTCGATGCCG  TCCTGCCCGA  AACCTTCCTG  CCTGAAAATC
     701  AGGATGGCGA  AGTGGCGGGT  TTTGAGAAAA  TGGACATCGG  CGGTCTGTTG
     751  GATGCCATGT  TGTGCGGAAA  CATGATGCAC  GACGCGCAAC  TGGTTACGCT
     801  GGACGCGTTT  TGCCGTTACG  GTCTGATTGA  TGCCGCCCAT  CCGCTGTCCG
     851  AGTGGCTGGA  CGGCATACGT  TTATAG
```

This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

-255-

1 MPTVRFTESV SKQDLDALE WAKASYGAES CWKTLYLNL PLGNLSPEWV
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGG RLQHLALGWH CAGLLDGWRN
 101 ECFDLTDGGG NPLFTLERAA FRPFGLLSRA VHLNGLTESD GRWHFWIGRR
 151 SPHKAVDPNK LDNTAAGGVS GGEMPSEAVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLSRSVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 DAMLSGNMMH DAQLVTLDAF CRYGLIDAAH PLSEWLDGIR L*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

meningitidis:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAES					
orf105a	MPTVRFTESVSKHDLDALEWAKASYGAES					
				10	20	30
orf105.pep	120	130	140	150	160	170
	CWKTLYLNGXPLGNLSPEWVERVXKDWAGCSESSDGIFLNADGWPMGGRLQHLALGWH					
orf105a	CWKTLYLNLPLGNLSPEWAERVKKDWAGCSESSDGIFLNADGWPMGGRLQHLARIWK					
	40	50	60	70	80	90
orf105.pep	180	190	200	210	220	230
	CAGLLDGWRNECFDLTDGGGNPLFTLERAXRXPXGLLSRAVHLNGLTESDGRWHFWIGRR					
orf105a	EAGLLHGWRECFDLTDGGSNPLFALERAAFRPFGLLSRAVHLNGLVESDGRWHFWIGRR					
	100	110	120	130	140	150
orf105.pep	240	250	260	270	280	290
	SPHKAVDPNKLDNTXAGGVSSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQSLHSLSRSVS					
orf105a	SPHKAVDPKLDNTAAGGVSSGELPSETVCRESSEEAGLDKTLPLIRPVSQSLHSLSRPVS					
	160	170	180	190	200	210
orf105.pep	300	310				
	RGVHNEILYV FDAVLP					
orf105a	RGVHNEILYV FDAVLPETFLPENQDGEVAGFEKMDIGGLLAAMLSGNMMHDAQLVTLDAF					
	220	230	240	250	260	270

The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

1 ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAAACACG ACCTTGATGC
 51 CCTATTTCGAG TGGGCAAAGG CAAGTTACGG TCGGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ATCTGTGCGC GGAATGGGCG
 151 GAGCGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTAGACGCG
 201 CATTTCCTG AATGCGGACG GCTGGCCAGA TATGGGCAGA CGCTTGCAGC
 251 ACCTCGCCCG AATATGGAAA GAAGCGGGAC TGCTTCACGG CTGGCGCGAC
 301 GAGTGTTCG ACCTGACCGA CGGCGGCAGC AATCCCTTGT TCGCGCTCGA
 351 ACGCGCCGCT TTCCGTCGCT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
 451 AGTCCGCACA AAGCAGTCGA TCCCGACAAA CTCGACAATA CTGCCGCCGG
 501 CGGTGTTTCC AGCGGTGAAT TGCCGTCTGA AACCGTGTGT CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
 601 TCGCAGCTGC ACAGCCTGCG CCCCCTCAGC CGGGGTGTGC ACAATGAAAT
 651 CTTGTATGTA TTCGATGCCG TCCTGCCGGA AACCTTCCTG CCTGAAATC
 701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTG
 751 GCTGCCATGT TGTCGGGAAA CATGATGCAC GACGCGCAAC TGTTTACGCT
 801 GACGCGGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This encodes a protein having amino acid sequence <SEQ ID 414>:

1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNL PLGNLSPEWA
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGR RLQHLARIWK EAGLLHGWRD


```

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGLL
251 AAMLSGNMMH DAQLVTLD AF CRYGLIDAAH PLSEWLDGIR L*

```

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

```

10 orf105a.pep      10      20      30      40      50      60
    MPTVRFTESVSKHDLDALEWAKASYGAESCWKTLYLNGPLGNLSPEWAERVKKDWEAG
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10 orf105-1        10      20      30      40      50      60
    MPTVRFTESVSKQDLDALEWAKASYGAESCWKTLYLNGPLGNLSPEWVERVKKDWEAG

15 orf105a.pep      70      80      90      100     110     120
    CSESSDGIFLNADGWPDMGRRLOHLARIWKEAGLLHGWRDECFDLTDGGSNPLFALERAA
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
15 orf105-1        70      80      90      100     110     120
    CSESSDGIFLNADGWPDMGRRLOHLALGWHCAGLLDGWRNECFDLTDGGGNPLFTLERAA

20 orf105a.pep      130     140     150     160     170     180
    FRPFGLLSRAVHLNGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
20 orf105-1        130     140     150     160     170     180
    FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPNKLNTAAGGVSSGEMPSEAVC

25 orf105a.pep      190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLHSLRPVSRGVHNEILYVDAVLPETFLPENQDGEVAG
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
25 orf105-1        190     200     210     220     230     240
    RESSEEAGLDKTLPLIRPVSQHLHSLRVSQHLHNEILYVDAVLPETFLPENQDGEVAG

30 orf105a.pep      250     260     270     280     290
    FEKMDIGLLAAMLSGNMMHDAQLVTLD ACRYGLIDAAHPLSEWLDGIRLX
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
30 orf105-1        250     260     270     280     290
    FEKMDIGLLDAMLSGNMMHDAQLVTLD ACRYGLIDAAHPLSEWLDGIRLX

35 orf105a.pep      250     260     270     280     290
    FEKMDIGLLAAMLSGNMMHDAQLVTLD ACRYGLIDAAHPLSEWLDGIRLX
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
35 orf105-1        250     260     270     280     290
    FEKMDIGLLDAMLSGNMMHDAQLVTLD ACRYGLIDAAHPLSEWLDGIRLX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N.*

gonorrhoeae:

```

40 orf105.pep      MVARRAHNPKVVGSNPX PATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 orf105ng        MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER 55

45 orf105.pep      QTAVCLRLQIQAVWLQSSALSRKPTMPTVRFTESVSKQDLDALEWAKASYGAESCWK 120
    |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
45 orf105ng        QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRFTESVSKQDLDALEWAKASYGAESCWK 115

50 orf105.pep      LYLNGLPLGNLSPEWVERVKKDWEAGCSESSDGIFLNADGWPDMGRRLOHLALGWHCAGL 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 orf105ng        LYLNRLPLGNLSPEWAERIKKDWAGCSESSNGIFLNADGWPDMGRRLOHLARTWNKAGL 175

55 orf105.pep      LDGWRNECFDLTDGGGNPLFTLERAXRXPGLLSRAVHLNGLTESDGRWHFWIGRRSPHK 240
    | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 orf105ng        LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHLNGLVESNGRWHFWIGRRSPHK 235

60 orf105.pep      AVDPNKLNTXAGGVSGGEMPSEAVCRESEEAGLDKTLPLIRPVSQHLHSLRVSQGVH 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 orf105ng        AVDPKLDNIAGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLRPVSQGVH 295

orf105.pep      NEILYVDAVLP 312
    |||:|||||:
orf105ng        NEILYVDAVLPETFLPENQDGEVAGFEKMDIGLLDAMLSKNMMHDAQLVTLD ACRYG 355

```

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

-257-

5
1 MVARRAHNPK VVGSNPAPAT KYQTPRFNAE GVLFFLPAA SVFCRIFLPA
51 AISERQAAVC LRLQIQAVWL QSSALCSRPK AMPTVRFTES VSKQDLDALE
101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWAE GCSSESSNGIF
151 LNADGWPDGM GRLQHLARTW NKAGLLHGWR NECFDLTDGG GNPLFTLERA
201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
251 SGGEMPSEAV CRESSEEAGL DKTFLPLIRP VSRLHSLRPV SRGVHNEILY
301 VFDAVLPETF LPENQDGEVA GFEMDIGGL LDAMLSKNMM HDAQLVTLDA
351 FYRYGLIDAA HPLSEWLDGI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10
1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
51 CCTGTTCCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTCGCC GGAATGGGCT
151 ACGGCATCA AAAAAGACTG GGAGGCAGGC TGCTCCGAGT CTCAGACGG
201 CATTMTTCTG AATGCGGACG GCTGGCCGGA TATGGGCGGA CGCTGCAGC
15
251 ACCTCGCCCG CACATGGAAC AAGGCGGGGC TGCTTACGG ATGGCGCAAC
301 GAGTGTTCG ACCTGACCGA CGGCGCGGGC AACCCTTGT TCACGCTCGA
351 ACGCGCCGCT TTCCGTCCGT TCGGACTACT CAGCCGCGCC GTCCATCTCA
401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC
451 AGTCCGCACA AAGCAGTCGa tCCGGCAAG CTCGACAATA TTGCCGGCGG
20
501 CGGTGTTTCC GCGGCGGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA
601 TCGCGGCTGC ACAGCCTTCG CCCCGTCAGC CGAGGTGTGC ACAATGAAAT
651 CCTGTATGTG TTCGATGCGG TCCTGCCCGA AACCTTCCTG CCGGAAATC
701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG
25
751 GATGCCATGT GTCGCAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

30
1 MPTVRFTESV SKQDLDALE RAKASYGAES CWKTLYLNR LPLGNLSPEWA
51 ERIKKDWEAG CSESSDGIFL NADGWPDMMG RLQHLARTWN KAGLLHGWRN
101 ECFDLTDGGG NPLFTLERA FRPFGLLSRA VHLNGLVESN GRWHFWIGRR
151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEEAGLD KTLFPLIRPV
201 SRLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
	orf105-1.pep	MPTVRFTESVSKQDLDALE	FEWAKASYGAES	CWKTLYLNLPLGNLSPEW	VERVKDWEAG		
	orf105ng-1	MPTVRFTESVSKQDLDALE	FERAKASYGAES	CWKTLYLNLPLGNLSPEW	AERIKKDWAE		
40		10	20	30	40	50	60
	orf105-1.pep	CSESSDGIFLNADGWPDMMG	RQLQHLALGWHCAGLLD	GWNECFDLTDGGGNPLFTL	ERAA		
45	orf105ng-1	CSESSDGIFLNADGWPDMMG	RQLQHLARTWNKAGLLHGWR	NECFDLTDGGGNPLFTL	ERAA		
		70	80	90	100	110	120
	orf105-1.pep	FRPFGLLSRAVHLNGLTES	DGRWHFWIGRRSPHKAVDP	PNKLDNTAAGGVSGGEMP	SEAVC		
50	orf105ng-1	FRPFGLLSRAVHLNGLVES	NGRWHFWIGRRSPHKAVDP	PGKLDNIAGGGVSGGEMP	SEAVC		
		130	140	150	160	170	180
	orf105-1.pep	RESSEEAGLDKTLPLIRPV	SQLHSLRSVSRGVHNEILY	VFDAVLPETFLPENQDGE	VAG		
55	orf105ng-1	RESSEEAGLDKTLPLIRPV	SRLHSLRPVSRGVHNEILY	VFDAVLPETFLPENQDGE	VAG		
		190	200	210	220	230	240
	orf105-1.pep	FEKMDIGGLLDAMLSGNMMH	DAQLVTLDAFCRYGLIDAAH	PLSEWLDGIRLX			
60	orf105ng-1	FEKMDIGGLLDAMLSKNMMH	DAQLVTLDAFYRYGLIDAAH	PLSEWLDGIRLX			
		250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5  sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
   >gi|1076928|pir|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
   (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
   [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
   pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
   Score = 105 bits (259), Expect = 4e-22
   Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)

10 Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAAPFRPFGLLSRAVHLNGLVESNGRW--HFWI 441
    N G+ WRNE + + P+ +ER F FG LS VH + + W+
   Sbjct: 96 NTFGIADQWRNELYTVYGKSKKPVLAVERGGFWLFGFLSTGVHCTMYIPATKEHPLRIW 155

15 Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRSLHSLR 621
    RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
   Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNL-PCGTVSYIK 214

20 Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGGLLDAMLSKNMMHDAQLV 798
    R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
   Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPIRINDGEVAGFSLLPLNQVLHELELKSFKPNCALVL 274

   Query: 799 LDAFYRYGLIDAAHP 843
    LD R+G+I HP
   Sbjct: 275 LDFLIRHGIITPQHP 289

```

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 419>:

```

30 1 ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
   51 CCAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
   101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
   151 TTGATATTTG GTAACATATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
35 201 ACCTGCATCG GCGGTAATCA GGTGTATGC ACCGgATACG rGkACAATTA
   251 CAGCGAAATT CGTGGAAGAT GGmsAAAAGG TTAAGGCTGG CGACAAGCTA
   301 TTTGCGCTTT CGACCTCACG TTTGCGGCGCA GGAGGTAGCG TGCAGCAGCA
   351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGCCAGAA CAGGAAGTGG
   401 GTCGTCTGAA GCTGATACAC GGGAATGAAA CGCGCagCcT TAAAGCAACT
40 451 GTCGAACGTT TGGAAAACCA GGAATCCAT ATTTCGCAAC AGATAGACGG
   501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
   551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45 1 MNRPKQPFFR PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
   51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
   101 FALSTSRFGA GGSVQQQLKT EAVLKKTAE QELGRLKLIH GNETRSLKAT
   151 VERLENQELH ISQQIDGQKR RIRLAEEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

-259-

orf107.pep MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT
 orf107a MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT
 10 10 20 30 40 50 60
 5
 orf107.pep TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT
 orf107a TVEGQILPASGVIRVYAPDXTITAKFXEDGEKVKAGDKLFALSTSRFGAGDSVQQQLKT
 10 70 80 90 100 110 120
 15
 orf107.pep EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
 orf107a EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO
 130 140 150 160 170 180
 20
 orf107.pep 189
 KYRFLSXQX
 orf107a KYRFLSANDAVPKQEMMNKAEELLEQKAKLDAYRREEVGLLQEIQTQNLTLXSLPQAA
 190 200 210 220 230

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

1 ATGAATAGAC CCAAGCAACC NTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
 25 51 CCAAACAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
 101 CCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
 151 TTGATATTG GTAACATAC GCGAAAGACA ACAGTGGAGG GACAAATTTT
 201 ACCTGCATCG GCGTAATCA GGGTGTATGC ACCGGATACG GGGACAATTA
 251 CNGCGAAATT CNTGGAAGAT GGAGAAAAGG TTAAGGCTGG CGACAAGCTA
 30 301 TTTGCGCTTT CGACCTCAGC TTTCCGGCGCA GGAGATAGCG TGCAGCAGCA
 351 GTTGAAAACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAAGTGG
 401 GTCGTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCCT TAAAGCAACT
 451 GTCGAACGTT TGGAACCA GGAAGTCCAT ATTTCGCAAC AGATAGACGG
 501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCGA AAATATCGTT
 35 551 TCCTATCCGC CAATGATGCA GTGCCAAAAC AAGAAATGAT GAATGTCAAG
 601 GCAGAGCTTT TAGAGCAGAA AGCCAACTT GATGCCTACC GCCGAGAAGA
 651 AGTCGGGCTG CTTCAGGAAA TCCGCACGCA GAATCTGACA TTGGNNAGCC
 701 TCCCCAAGC GGCATGA

This encodes a protein having amino acid sequence <SEQ ID 422>:

40 1 MNRPKQFFR PEVAVARQTS LTGKVILTRP LSFLWTTFA SISALLIILF
 51 LIFGNYTRKT TVEGQILPAS GVIRVYAPDT GTITAKFXED GEKVKAGDKL
 101 FALSTSRFGA GDSVQQQLKT EAVLKKTAE QELGRLKLIH GNETRSLKAT
 151 VERLENQELH ISQQIDGQKR RIRLAEEMLO KYRFLSANDA VPKQEMMNK
 201 AEELLEQAKL DAYRREEVGL LQEIQTQNL TLXSLPQAA*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.*

gonorrhoeae:

50 orf107.pep MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT 60
 orf107ng MNRPKQFFRPEVAIARQTSLTGKVILTRPLSFSLWTTTFASISALLIILFLIFGNYTRKT 60
 orf107.pep TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT 120
 55 orf107ng TMEGQILPASGVIRVYAPDXTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT 120
 orf107.pep EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO 180
 orf107ng EAVLKKTAEQELGRLKLIHENETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLR 180
 60 orf107.pep KYRFLSXQ 188
 orf107ng KYRFLSAQ 188

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQFFFR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLILF
51 LIFGNYTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLAE QELGRLKLIH ENETRSLKAT
151 VERLENQKLH ISQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
51 GTGCGGCAAA TCCGTAATA CGGCGGTACA GCCGCAAAAC GCGGTACAAA
15 101 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCAG AAAACGGCGT
20 351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25 1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
51 AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNI A EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30 1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCGC
51 CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
35 201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GGCGACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCCGTGCAG AACGGACGCT ATGTGCTGGA
40 501 AATCGACAGC GAAGGGGCGT TTTATTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45 1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51 AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNI A EDGGKLT DYL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

5      orf108.pep  MLNTFFAVLGGCLLLPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      orf108ng    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
10     orf108ng    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

      orf108.pep  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
15     orf108ng    LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

      orf108-1.pep MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
20     orf108ng-1  MLKIPFAVLGGCLLLAACGKSENTAEQPQNAASAPKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108-1.pep GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng-1  GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

25     orf108-1.pep LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
      orf108ng-1  LFAKLVGNI AEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

30      1  ATGCTGAAAA  tacctTTTGC  CGTGTtgggc  ggCtgccctGC  TGCTTGCCGC
      51  CTGCGGCAAA  TCCGAAAATa  cggcgggAAC  GCCGCAAAAT  gcggCACAAA
      101  GCGCGCCGAA  ACCGGTTTTC  AAAGTCAAAT  ACATCGACAA  TACGGCGATT
      151  GCCGGTTTGG  CTTTGGGACA  AAGTAGCGAA  GGCAAAACCA  acgacgGCAA
      201  AAAACAAATC  AGTTATccgA  TTAAAGGCTT  GCCGGAACAA  Aacgcgctcc
      251  gGCTGACCGG  AAAGCATCCC  AACGACTTGG  AagccgtcgT  CGGCAAATGT
35     301  ATGGAACCG  ACGGAAAGGA  CGCGCCTTCG  GGCTGGGCGG  AAAACGGCGT
      351  GTGCCATACC  TTGTTTGCCA  AACTGGTGGG  CAATATCGCC  GAAGACGGCG
      401  GCAAACGAC  TGATTACCTG  ATTCGCATT  CCGCCCTGCA  ACCCTATCAG
      451  GCAGGCAAAA  GCGGCTATGC  CGCCGTGCAG  AACGGACGCT  ATGTGCTGGA
      501  AATCGACAGC  GagggGCGT  TTTATttccg  ccgccgccat  tattgA

```

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

      1  MLKIPFAVLG  GCLLLAACGK  SENTAEPQN  AAQSAPKPVF  KVKYIDNTAI
      51  AGLALGQSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
      101  METDGKDAPS  GWAENGVCHT  LFAKLVGNI  A  EDGGKLT DYLV  ISHSALQPYQ
      151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*

```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

-262-

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
 51 CGgATTATC GATgcatg cGgCGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
 151 CTGCAAGCAG CCGCTGCTAC GTTTTAGCT ACGGTTTCTT TTGCACGCAA
 5 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 251 TAGGCGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTgCTgGCGG TCGTGCCGGT TTTGTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 10 401 TTTTCTGTT cGGGCTGACG GTCGC.ACCG CTTTGGGTT TTTACGACGG
 451 TGTGTTCGGA CCGGGTGTG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
 501 TGCTCGGCTG CAAgCTGTTG AACCGATGT CTTACACCAA ATTGGCGAAC
 551 GTTGCCTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT
 601 TATTTTCCCG ATTGCGGCAA CGaTGGCGGT CGGTGCGTTT GTCGGtCGCA
 651 ATTTAgGTGC GAGATTTGCC GTaCgctTCG GTTCGAAGCT GATTAA

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGLRR
 151 CVRTGCRILVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD
 20 201 YFPDCGNDGG RVCRCRCEFR EICRTLRFEEA D*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTCGGTTG CGATGATTGC
 51 CGGATTATC GATGCGATTG CGGGCGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
 151 CTGCAAGCAG CCGCTGCTAC GTTTTAGCT ACGGTTTCTT TTGCACGCAA
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 251 TAGGCGCGT GGCCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTGCTGGCGG TCGTGCCGGT TTTGTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 10 401 TTTTCTGTT CCGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
 451 GTTTCGGAC CGGGTGTGCG CTCGTTTTT CTGATTGCCT TTATTGTTTT
 501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA ATTGGCGAAC
 551 TTGCCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGAT
 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
 35 651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
 701 TGCTGATTGT CATCAGCATT TCGATGCTG TGAAATGTG GATAGACGAG
 751 AGAAATCCCG TGTATCAGAT GATTGTTTCG ATGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 40 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
 151 VFGPGVSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS
 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
 251 RNPLYQMIVS MF*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
50	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
		10	20	30	40	50	60
55	orf109.pep	TVSFARKGLIDWKKGPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109a	TVSFARKGLIDWKKGPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120

```

      130      140      150      160      170      180
orf109.pep KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ
5 orf109a   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

      1 ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
    51 CCGATTTATC GATGCGATTG CCGGTGGGGG TGGTTTGATT ACGCTGCCTG
10 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
    151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
    201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
    251 CAGGCGGCGT GGTCTGGTCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
    301 CTGCTGGCGG TCGTGGCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
15 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
    401 TTTTCTGTGT CCGTCTGACG GTTGCAACAC TTTTGGGTTT TTACGACGGT
    451 GTGTGCGGAC CGGGTGTCTCG CTCGTTTTTT CTGATTGCCT TTATGTTTT
    501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
    551 TTGCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CGGTTCGATT
20 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
    651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
    701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAAATTGTT GATAGACGAG
    751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25      1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
    51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
10 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
    151 VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLLHGS I
30 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
    251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

      10      20      30      40      50      60
orf109a.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60

      70      80      90      100     110     120
orf109a.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
40 orf109-1  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100     110     120

      130     140     150     160     170     180
orf109a.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
45 orf109-1  KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130     140     150     160     170     180

      190     200     210     220     230     240
orf109a.pep LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
50 orf109-1  LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190     200     210     220     230     240

      250     260
orf109a.pep SMAVKLLIDERNPLYQMIVSMFX
55 orf109-1  SMAVKLLIDERNPLYQMIVSMFX
      250     260
60

```


Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N.gonorrhoeae*:

5	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
	orf109ng	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA	60
10	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP	120
	orf109ng	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP	120
15	orf109.pep	KLDGSKEGKARMSFFLFGLTVXTAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ	180
	orf109ng	KLDGSKEGKARMSFFLFGLTVATAFGLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ	180
	orf109.pep	IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRTLRFAD	231
	orf109ng	IGERCLQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFAD	231

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

1	MEDLYIILAL	GLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNK
51	LQAAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA	LSVSLVSKDI
101	LLAVVPVLLI	FVALYFVFSP	KLDGSKEGKA	RMSFFLFGLT	VATAFGFLRR
151	CVRTGCRLVF	SDCLYCFARL	QAVERDVLHQ	IGERCLQSWF	ATGIPAARFD
201	YFPDCGNDGG	RCVCRCEFR	EICRPLRFEA	D*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

1	ATGGAAGATT	TATACATAAT	ACTCGCTTTG	GGTTTGGTTG	CGATGATCGC
51	CGGATTATATC	GATGCGATTG	CGGGCGGGGG	TGGTTTGATT	ACGCTGCCTG
101	CACCTCTGTT	GGCAGGTATT	CCTCCCGTGT	CGGCAATTGC	CACCAACAAG
151	CTGCAAGCAG	CCGCTGCTAC	GTTTTTCGGCT	ACGGTTTCTT	TGACACGCAA
201	AGGTTTGATT	GATTGGAAGA	AAGGTCTCCC	GATTGCCGCA	GCATCGTTTG
251	CAGGCGGCGT	GGTCGGTGCA	TTATCGGTCA	GCTTGGTTTC	CAAAGATATT
301	TTGCTGGCGG	TCGTGCCGGT	TTTGTGATA	TTTGTGCGC	TGATTTTGT
351	GTTTTCGCCC	AAGCTCGACG	GCAGTAAGGA	AGGCAAAGCC	AGAATGTCTT
401	TTTTTCTATT	CGGGCTGACG	GTTGCACCGC	TTTGGGTTT	TTACGACGGT
451	GTGTTCCGAC	CGGGTGTCGG	CTCGTTTTTT	CTGATTGCCT	TTATTGTTTT
501	GCTCGGCTGC	AAGCTGTTGA	ACGCGATGTC	TTACACCAAA	TGGCGAACG
551	TTGCTTGCAA	TCTTGGTTTC	CTATCGGTAT	TCCTGCTGCA	CGGTTTCGATT
601	ATTTTCCCGA	TTGTGCAAC	GATGGCGGTC	GGTGCGTTTG	TCGGTGCGAA
651	TTTAGTGCG	AGATTGCGG	TCCGCTTCGG	TTCGAAGCTG	ATTAAGCCGC
701	TGCTGATTGT	CATCAGCATT	TCGATGGCTG	TGAAATTGTT	GATAGACGAG
751	AGAAATCCGC	TGTATCAGAT	GATTGTTTCG	ATGTTTTAA	

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

1	MEDLYIILAL	GLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNK
51	LQAAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA	LSVSLVSKDI
101	LLAVVPVLLI	FVALYFVFSP	KLDGSKEGKA	RMSFFLFGLT	VAPLLGFYDG
151	VFGPGVGSFF	LIAFIVLLGC	KLLNAMS YTK	LANVACNLGS	LSVFLHGS I
201	IFPIVATMAV	GAFVGANLGA	RAVRFGSKL	IKPLLIVISI	SMAVKLLID E
251	RNPLYQMIVS	MF*			

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

		10	20	30	40	50	60
orf109ng-1.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA						
orf109-1	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA						
		10	20	30	40	50	60
		70	80	90	100	110	120
orf109ng-1.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP						

-265-

```

      |||:|||||
orf109-1 TVSFARKGLIDWKKGLPIAAASVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100     110     120
5
      130     140     150     160     170     180
orf109ng-1.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      |||:|||||
orf109-1 KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMS YTK
      130     140     150     160     170     180
10
      190     200     210     220     230     240
orf109ng-1.pep LANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGARFAVRFGSKLIKPLLVISI
      |||:|||||
orf109-1 LANVACNLGSLSVFLLHGSIIFFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLVISI
      190     200     210     220     230     240
15
      250     260
orf109ng-1.pep SMAVKLLIDERNPLYQMIVSMFX
      |||:|||||
20
orf109-1 SMAVKLLIDERNPLYQMIVSMFX
      250     260

```

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

25 sp|P29942|YCB9_PSEDE_HYPOTHETICAL_27.4_KD_PROTEIN_IN_COBO_3'REGION_(ORF9)
>gi|94984|pir|I138164_hypothetical_protein_9_-_Pseudomonas_sp_>gi|551929
(M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

30 Query: 41 PPVSAIATNKLOXXXXXXXXXXXXXKRLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKDI 100
PP+ + TNKLQ R+G ++ K+ LP+ D+
Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFF 160
L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
35 Sbjct: 103 LKAILPFLIIAIALYFGLKPNM-GDVDQHSRVTPEVFTLTPLVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMS YTKLANVACNLGSLSVFLLHGSIIFFPIVATMAVGAFVGANLGA 220
++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
40 Sbjct: 162 MLGFVTLAGFGVLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLMLGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLVISISMVAVKLLIDERNPL 254
R+A+ G+K+IKPLL+++SI++A++LL D +PL
Sbjct: 222 RYAMAKGAKIIPLLVIVSIALAIRLLADPTHTPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

```

50      1 ..CTGCTAGGGT ATTGCATCGG TTATCGGTAC GGCTGTTGCA GCAAAACCAG
51      CCGCAGACGG ATTATTTGGT CAAATTCGGA TCGTTTTGGG CGAG.ATTTT
101     TGGTTTTCTG GGACTGTATG ACGTCTATGC TTCGGCATGG TTTGTCGTTA
151     TCATGATGTT TTTGGTGGTT TCTACCACTT TGTGCCTGAT TCGCAATGTG
201     CCGCCGTTCT GCGCGGAAAT GAAGTCTTTT CCGGAAAAGG TTAAAGAAAA
55     251 ATCTCTGGCG GCGATGCGCC ATTCTTCGCT GTTGATGTA AAAATTGCGC
301     CCGAGGTTGC CAAACGTTAT CTGGAAGTAC AAGGTTTTCA GGGGAAAACC
351     ATTAACCGTG AAGACGGGTC GGTCTGATT GCCGCCAAA AAGGCACAA
401     GAACAAATGG GGCTATATCT TTGCCCATGT TGCTTTGATT GTCATTTGCC
451     TGGCGGGTT GATAGACAGT AACCTGCTGT TGAACTGGG TATGCTGACC
60     501 GGTCGGATTG TTCCGGACAA TCAGGCGGTT TATGCCAAGG ATTTC.AAGC

```

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA
601 TATTTCCG.A GGGGCAGAgT GCGGATGTGG TTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5 1 ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
 51 MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
 101 EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKGYIF AHVALIVICL
 151 GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI*GQRQY
 201 FXRGRVRMWF S*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA					
15	orf110			LLGIASVIGTLLQONQPQTDYLVKFGSFWA			
				10	20	30	
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
	orf110	XIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH					
		40	50	60	70	80	90
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGYIFAHVALIVICL					
	orf110	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGYIFAHVALIVICL					
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLLKLGMILTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVVF					
	orf110	GGLIDSNLLLKLGMILTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF					
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLPFEVKKLKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT					
	orf110	SX					

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep	LLGIASVIGTLLQONQPQTDYLVKFGSFWA	30
	orf110ng	MSKSRISPTLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGPFWT	60
50	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	90
	orf110ng	RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH	120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSVLIAAKKGTMNKGYIFAHVALIVICL	150
55	orf110ng	SSLLDVKIAPEVAKRYLEVGRFGQKTVSREDGSVLIAAKKGTMNKGYIXAHVALIVICL	180

```

orf110.pep  GGLIDSNLLKLGLMTGRIFRTIRRFMPRIXKPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
| | | : | | | | | | | : | | | : | | | | | | | : | | | | | | | | | : | | | |
orf110ng    GRLINXNLLKLGLLAGSIFRNNRRVMPRIKPEIWWGGVQSLIKGQRQYFQRGKVRMWF  240
5   orf110.pep  S  211
      |
      orf110ng  S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10   1  MSKSRIPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
      51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVVSTSL CLIRNVPPFW
      101  REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVR GFQGTVSRE
      151  DGSVLIAAKK GTMNKGWYIX AHVALIVICL GRLINXNLLL KLGMLAGSIF
      201  RNNRRVMPRI SKPESIWWGV QSLIKGQRQY FQRGKVRMWF S*

```

15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20   1  ATGCCGCTCG AAACACGCCT GCCGAACCTT ATCCGCGTCT TGATATTTGC
      51  CCTGGGTTTC ATCTTCCTGA ACGCCTGTTC GGAACAAACC GCGCAAACCG
      101  TTACCCTGCA AGGCGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
      151  TCAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAAACGCAT
      201  CGATGACGCG CTAAAGAAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
25   251  ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCTCCGC
      301  ATTTCAAGCG ACTTCGCACA CGTFACTGCC GAAGCCGTCC GCCTGAACCG
      351  CCTGACACAC GCGCGCTGG ACGTAACCGT CGGCCCTTG GTCAACCTTT
      401  GGGGATTCGG CCCGACAAA TCCGTATCCC GTGAACCGTC GCCGGAACAA
      451  ATCAAAACAG CGGCATCTTA TACGGGCATA GACAAATCA TTTGAAACA
30   501  AGGCAAAGAT TACGCTTCCT TGAGCAAAAC CCACCCCAAG GCCTATTTGG
      551  ATTTATCTTC GATTGCCAAA GGCTTCGCGC TTGATAAAGT TCGGGCGCAA
      601  CTGGA AAAAT ACGGCATTC AATTATCTG GTCGAAATCG GCGGCGAGTT
      651  GCACGGCAAA GGCAAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
      701  AGCAGCCCAA TATCGTCCAA GCGGCAATA CGCAGATTAT CGTCCCGCTG
35   751  AACAAACGTT CGCTGCCAC TTCGCGGAT TACCGTATTT TCCACGTCGA
      801  TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
      851  CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGAGA CAGTGCGATG
      901  ACGGCGGACG GCTGTCCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
      951  CTAAAGCTG GCAGAGCGCG AAAAATCGC TGTTCCTG ATTGTCAGGG
40   1001  ATAAAGCGG CTACGCACC GCCATGTCTT CCGAATTGA AAAACTGCTC
      1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45   1  MPSETRLPNF IRVLIFALGF IFLNACSEQT AQTVTLOGET MGTYYTVKYL
      51  SNNRDKLPSP AEIQKRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR
      101  ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWFGPDK SVTREPSPEQ
      151  IKQAASYTGI DKILKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
      201  LEKYGIQNYL VEIGGELHKG GKNARGEPRW IGIEQPNIVQ GGNTQIIIVL
      251  NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVADSAM
      301  TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
50   351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351 aa overlap with an ORF (ORF111a) from strain A of *N.*

meningitidis:

5	orf111a.pep	10	20	30	40	50	60
	orf111	10	20	30	40	50	60
10	orf111a.pep	70	80	90	100	110	120
	orf111	70	80	90	100	110	120
15	orf111a.pep	130	140	150	160	170	180
	orf111	130	140	150	160	170	180
20	orf111a.pep	190	200	210	220	230	240
	orf111	190	200	210	220	230	240
25	orf111a.pep	250	260	270	280	290	300
	orf111	250	260	270	280	290	300
30	orf111a.pep	310	320	330	340	350	
	orf111	310	320	330	340	350	

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCG	AAACACGCCT	GCCGAACCTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTT	ATCTTCCTGA	ACGCCTGTTC	GGAACAAACC	GCGCAAACCG
	101	TTACCCCTGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCNTCACCT	GCCGAAATAC	AAAANCGCAT
	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
45	251	ACTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCTCCGC
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAACCGT	CGGCCCTTG	GTCAACCTTT
	401	GGGGATTCCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAAG	CAGCATCTTA	TACGGGCATA	GACAAAATCA	TTTGAAACA
50	501	AGGCAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAATAAT	ACGGCATTCA	AAATTATCTG	GTGAAATCG	GCGNGAGTT
	651	GCACGGCAAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATTT	TCCACGTCGA
	801	TAAAGCGGC	AAACGCCTCT	CCCATATCAT	TAATCCGAAC	AACAAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCGATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAAACTCGC	TGTTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTTGA	AAAACGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

1 MPSETRLPNF IRTLIFALSF IFLNACSEQT AQTVTLQGET MGTTYTVKYL
51 SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR

-269-

101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPL VNLWFGPDK SVTREPSPEQ
 151 IKQAASYTGI DKIIILKQKGD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
 201 LEKYGIQNYL VEIGGELHGK XKNARGEPR IGIEQPNIVQ GGNTQIIIVPL
 251 NNRSXATSGD YRIFHVDKSG KRLSHIINPN NKRPISHNLA SISVXADSAM
 5 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
 351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
	orf111ng	MPSETRLPNLIRALIFALGFIFLNACSEQTAQVTTLQGETMGTTYTVKYLNNRDKLPSP					
	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQVTTLQGETMGTTYTVKYLNNRDKLPSP					
15		10	20	30	40	50	60
	orf111	AKIQKRIDDALKEVNROMSTYQTDSEISRFNQHTAGKPLRISDFAHVTA EAVRLNRLTH					
		:					
20	orf111	AEIQKRIDDALKEVNROMSTYQPDSEISRFNQHTAGKPLRISDFAHVTA EAVRLNRLTH					
		70	80	90	100	110	120
	orf111ng	GALDVTVGPLVNLWFGPDKSVTREPSPEQIKQAASYTGIDKIIILQOGKDYASLSKTHPK					
25	orf111	GALDVTVGPLVNLWFGPDKSVTREPSPEQIKQAASYTGIDKIIILQOGKDYASLSKTHPK					
		130	140	150	160	170	180
	orf111ng	AYLDLSSIAKGFVGDKVAGELEKYGIQNYLVEIGGELHGKGNHAGEPWRIGIEQPNIIQ					
30	orf111	AYLDLSSIAKGFVGDKVAGELEKYGIQNYLVEIGGELHGKGNARGEPRIGIEQPNIVQ					
		190	200	210	220	230	240
	orf111ng	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLASISVVSDSAM					
35	orf111	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPISHNLASISVVADSAM					
		250	260	270	280	290	300
	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKLLRX					
40	orf111	TADGLSTGLFVLGETEALRLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
		310	320	330	340	350	
45	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKLLRX					
	orf111	TADGLSTGLFVLGETEALRLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
		310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

	1	ATGCCGCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTC	ATCTTCCTGA	ACGCCTGTTC	GGaacaacC	CGCGAaacg
50	101	TTACCCTGCA	AGGCGAAacg	aTGGGTACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACAAACT	CCCCTCCCT	GCCAAAATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACCAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCTCCGC
	301	ATTTCAAGCG	ATTTCCGACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
55	401	GGGGGTTCGG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTGTCAACA
	501	AGGCAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
	601	CTGGA AAAAT	ACGGCATTCA	AAATTATCTG	GTGAAAatcg	gcggcGAGTT
60	651	GCACGGGAAA	GGCAAAAATG	CGCACGGCGA	ACCGTGGCGC	ATCGGTATAG
	701	AGCAACCCAA	TATCATCCAA	GcgGCAata	CGCAGATTat	cgteccgctg
	751	aaCaaccgtt	cgctTGCCAC	TTCCGGCGAT	TaccgtaTTT	tccacgtcgA
	801	TAAAAacggc	aaacgccttt	cccacaTCAT	CAATCCCaAC	aacAAACgac
	851	ccATCAGcca	caacctgccc	tccatcagcg	tggtctcAGA	CAGTGCAATG
65	901	ACGGCGGACG	GTTtatCCAC	AGGATTATTT	GTTTTAGGCG	AAACCGAAGC
	951	CTTAAGGCTG	GCAGAACAAAG	AAAAACTCGC	TGTTTTCTTA	ATTGTCCGGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTGC CAAGCTGCTC
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5 1 MPSETRLENL IRALIFALGF IFLNACSEQT AQTVTLOGET MGTITYTVKYL
51 51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS YQTDSEISRF NQHTAGKPLR
101 101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFPGDK SVTREPSPEQ
151 151 IKQAASYTGI DKIIQQQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
201 201 LEKYGIQNYL VEIGGELHGK GKNAHGEFWR IGIEQPNIIQ GGNTQIIIVPL
251 251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVSAMS
10 301 TADGLSTGLF VLGETEALRL AEQEKLAFL IVRDKDGYRT AMSSEFAKLL
351 351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

15 sp|P44550|YOJL HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4
hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)
>gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
Score = 353 bits (896), Expect = 9e-97
Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)
20 Query: 7 LPNLIRALIFALGFIFLNACSEQT AQTVTLOGETMGTITYTVKYL SNNRDKLPSPAKIQKR 66
+ LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +
Sbjct: 1 MKKLISGIIAVAMASLAACQKET-KVISLSGKTMGTTYHVKYLDGSGITATSE-KTHEE 58
25 Query: 67 IDDALKEVNRQMSYQTDSEISRFNQHT-AGKPLRISSDFAHVTA EAVRLNRLTHGALDV 125
I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV
Sbjct: 59 IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTEGALDV 118
30 Query: 126 TVGPLVNLWGFPGDKSVTREPSPEQIKQAASYTGIDKIIQQQKDYASLSKTHPKAYLDL 185
TVGP+VNLWGFPG+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL
Sbjct: 119 TVGPPVNLWGFPGPEKREKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178
35 Query: 186 SSIAGFGVDKVAGELEKYGIQNYLVEIGGELHGK GKNAHGEFWRIGIEQPNIIQGGNTQ 245
SSIAGFGVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +
Sbjct: 179 SSIAGFGVDQVAEKLEQLNAQNYMVEIGGEIRAKGKNIEGKWPQIAIEKPTTTGERAVE 238
Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVVSAMSADGL 305
++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL
Sbjct: 239 AVIGLNNMGMASSGDYRIY-FEENGKRAHEIDPKTGYPIQHHLASITVLAPTSMTADGL 297
40 Query: 306 STGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKL 349
STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL
Sbjct: 298 STGLFVLGEDKALEVAEKNLAVYLIIRTDNFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

50 1 ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCACG CGTCCCGTCA
51 51 AAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
101 101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTGCAAT CGGCGGCGAG
151 151 GTGTTGTAC GGCAAAATGA AGGCAGCCKA yTGGCAATCG GCGTGATGGG
201 201 CGGCAGGGCC GGCCAGCACG CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
251 251 gCAGTGATTT GTATGTTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
301 301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
351 351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
401 401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
55 451 GAAGGCATTG TCGGAAAAGG CAATAATGTG CCGTTTACC TACAACCGCA
501 501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CCGCTTACC GACAGCGAGG
551 551 GGACGCGGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
601 601 AtTCGGGCAA AAACCCGTTT TGCTTGGCGT AACGGTGTCA ATCTTCAGCC
651 651 TTTTGGCGCT TTTAATGTTt TGCACAGGTC AAAATCTTTC GCGTGGA
60 701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

-271-

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

1 ..PCRRQGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGVQIGGE
 51 VFVRQNEGSX LAIGVMGGRA QHASVNGKG GAAGSDLYGY GGGVYAAWHQ
 101 LRDKQTGAYL DGWLQYQRFK HRINDENRAE RYKTKGWTAS VEGGYNALVA
 151 EGIVGKGNV RFYLQPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
 201 IRAKTRFALR NGVNLPFAA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
 251 FGIEAGWKGH MSA..

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGVQIGGEVFVRQNEGSXLAI 63
 + D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I
 virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQVWQGTAPVEGYRKGVQLGGEVFTWQNESNQLSI 455
 15 Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH 121
 G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H
 virg-h 456 GLMGQAQERSTFHNPDNTNLTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQFRH 515
 20 Orf35 122 RINDENRAERYKTKGWTASVEGGYNALVAEGIVGKGNVRFYLQPAQFTYLGVNGGFTD 181
 RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPAQ TYLGVNG F+D
 virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKNLSLRVYLQPAQLTYLGVNGKFS 575
 25 Orf35 182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLPFAAFNVLHRSKSGVEMDGEKQTL 241
 SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEMDGE++ +
 virg-h 576 SENAHVNLLGSRQLQTRVGVQAKAQFSLYKNIAIEPFAAVNALYHNKPFVEMDGERVI 635
 Orf35 242 AGRTALEGRFGIEAGWKGHMS 262
 +TA+E + G+ K H++
 30 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.**meningitidis*:

35 orf35.pep 10 20 30
 PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIRG
 :||||| :||||| :|||||
 orf35a QRLAIPAEAEAVLYAQAYAAANTLFLRAADRGGDDVYAADPSRQKLWLRFIGGRSHQNIRG
 310 320 330 340 350 360
 40 orf35.pep 40 50 60 70 80 90
 GAAADGWRKGVQIGGEVFVRQNEGSXLAIGVMGGRAGQHASVNGKGAAGSDLYGYGGGV
 :||||| :||||| :||||| :||||| :||||| :|||||
 orf35a GAAADGRRKGVQIGGEVFVRQNEGSRLAIGVMGGRAGQHASVNGKGAAGSYLHGYGGGV
 370 380 390 400 410 420
 45 orf35.pep 100 110 120 130 140 150
 YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV
 :||||| :||||| :||||| :||||| :||||| :|||||
 50 orf35a YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV
 430 440 450 460 470 480
 55 orf35.pep 160 170 180 190 200 210
 GKGNVRFYLQPAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN
 :||||| :||||| :||||| :||||| :||||| :|||||
 orf35a GKGNVRFYLQPAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN
 490 500 510 520 530 540
 60 orf35.pep 220 230 240 250 260
 LQPFPAFNVLHRSKSGVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA
 :||||| :||||| :||||| :||||| :|||||

orf35a LQPF AAFNV LHRSKS FGVEMD GEKQTL AGR TALEGR FGI EAGWKGHMSARIGYGKRTDGD
550 560 570 580 590 600

orf35a KEAALSLKWLFX
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

```

1  ATGTTTCAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCCGA
51  CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT
101 ATTTTTCCAG CGGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC
151 GAAATCAATA TCCAAGGTAA AACTACAAT AGCGGCATAC TCGCCGTCGA
201 TAATATGCCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GGTAAAGAAG CAATTACAGG ATTTATACAA AACAAGACCC
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA
351 GCTTGGACCA AAATTTAGTA TACTCAAACA GAAAAACCCG GATTTAATTA
15  401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCACG GTCGCCGGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCTT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA
601 ACTTCCGATA ATGCCCGCAT CCGCCTGAAC ACGAAAGATG AAAAAGTAC
20  651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCCTGTTC GGCTACGACG
701 TCGGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGGCG TGGTTTTGGA ACGCCGGCCG GAAATCTGA AAACGCTCGA
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTTGCCT
851 TTAAACAAAA TTACCGGCAG GGAAGTGTACG AATTATTGCT CAAGCAATGC
25  901 GAAGGCGGAT TTGCTTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA
951 AGCGGTTTGA TATGCCCAAC AGGCTTATGC GGCAATACT TTGTTCCGGC
1001 TCGGTGCCGC CGACAGGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCCT
1051 CAAAATTGTT GGCTGCGCTT CATCGCGCGC CGGTGCGATC AAAATATACG
1101 GCGCGGCGCG GCTGCGGACG GCGCGCGCAA AGCGGTGCAA ATCGGCGCGC
30  1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CCGCGTGATG
1201 GCGGGCAGGG CTGGCCAGCA CGCATCAGTC AACGGCAAAG GCGGTGCGGC
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG TGTATTGCT GCGTGGCATC
1301 AGTTGCCGGA TAAACAAACG GGTGCGTATT TGGACGGCTG GTTGCAATAC
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA
35  1401 AACCAAAGGT TGGACGGCTT CTGTGCAAGG CCGCTACAAC GCGCTTGTGG
1451 CGGAAGGCGT TGTGCGAAAA GGCAATAATG TCGGTTTGA CCGTCAACCG
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GCGGCTTTA CCGACAGCGA
1551 GGGGACGCGG GTCGGACTGC TCGGCAGCGG TCAGTGGCAA AGCCGCGCCG
1601 GCATTCCGGC AAAAACCCTT TTTGCTTTGC GTAACGGTGT CAATCTTCAG
40  1651 CCTTTTCCCG CTTTAAATGT TTGTCACAGG TCAAATCTT TCGGCGTGGA
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC
1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACCCATCGGA
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATTGT CGTCAAATG
1851 GCTGTTTTGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1  MFRAQLGSNT RSTKIGDDAD FSFSDKPKPG TSHYFSSGKT DQNSSEYGYD
51  EINIQQKNYN SGILAVDNMP VVKYITDITY GDNLKDAVKK QLQDLYKTRP
101  EAWEEKKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
50  151  TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE
201  TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251  QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLLKQC
301  EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFLGLRAADRQ DDVYAADPSR
351  QKLWLRFIGG RSHQNIRGGA AADGRRKGVQ IGGEVFRQON EGSRLAIGVM
55  401  GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLOQ
451  QRFKHRINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLOP
501  QAQFTYLGVN GGFTDSEGTA VGLLGSGQWQ SRAGIRAKTR FALRNGVNLQ
551  PFAAFNVLHR SKSFGVEMDG EKQTLAGR TA LEGRFGIEAG WKGHMSARIG
601  YGKRTDGDKE AALSLKWLF*

```

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N. gonorrhoeae*:

```

orf35.pep                                PCRRQGGDDVYAAHASRQKLWLRFIGGRSHQNIRG 34
                                         :::|:: |::|::| |::|::|
orf35ngh                                FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370

```

10

15

20

25

30

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35

40

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50

55

```

1      ..GCAGTGTGCC  TnCCGATGCA  TGCACACGCC  TCAnATTTGG  CAAACGATTC
51     TTTTATCCGG    CAGGTTCTCG  ACCGTcAGCA  TTTcGAACCC  GACGGGAAAT
101    ACCACCTATT     CGGCAGCAGG  GGGGAACTTg  CCGAGcGCCA  GTCTCATATC
151    GGATTTGGGAA    AAATACAAG  CCATCAGTtG  GGCAACcTGA  GTATTCAACA
201    GCGCGCCATT     AAAGGAAATA  TCGGTCATAT  TGTCGGcTTT  TCCGATCAGC
251    GGCACGAAGT     CCATTCCCCs  TTCGACAACC  ATGCCTCACA  TTCCGATTCT
301    GTAGGAAGCCG    GTAGTCCCGT  TGACGGATtT  AGCCTTTTAC  CCATCCATTG
351    GGACGGATAC     GAACACCATC  CCGCGCAGCG  CTATGACGGG  CCACAGGGCG
401    GCGGCTATCC     CGCTCCCAAA  GCGCGAGAGG  ATATATACAG  TTACGACATA

```

-274-

5
 451 AAAGGCGTTG CCCAAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC
 501 CGGACAACGG CTTGCCGACC GTTTCACAA TGCCGGTAGT ATGCTGACGC
 551 AAGGAGTAGG CGACGGATTG AACGCGCCA CCCGATACAG CCCCAGCTG
 601 GACAGATCGG GCAATGCTCG CGAAGCCTTC AACGGCACTG CAGATATCGT
 651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

10
 1 ..AVCLPMHAHA SXLANDSFIR QVLDRQHFEF DGKYHLFGSR GELAERQSHI
 51 GLGKIQSHQL GNLMIQQAII KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
 101 DEAGSPVDGF SLYRIHWDGY EHPADGYDG PQGGGYAPAK GARDIYSYDI
 151 KGVAQNIRLN LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPDL
 201 DRSGNAAEAF NGTADIVKNI IGAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

20
 orf46.pep AEYVQFSIDLFSVGKSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 45
 orf46ng PKTGVPFDGKGFPNFEKHVKYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLNKLTTTR 217
 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105
 orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277
 orf46.pep RVIQQTSAADKHGXLSSDSGN 126
 orf46ng RVIQQTSAADKHGVLSSDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30
 1 ..RRLKHCCHAR LGSFHRKQD GAHQRFGRYG ATQRLCRSSH PRLGSPKQPC
 51 RTRHRSRQQY LYGSHPHQD WSCPGKIQLG RHHGTSCRAV ADXRDRICER
 101 EIRRQQRQXR CRLGKIPSLI IPKYPLKLEQ RYKGENITSS TVPPSNGKNV
 151 KLADQRHPKT GVPFDGKGFP NFEKHVKYDT KLDIQELSGG GIPKAKPVFD
 201 AKPRWEVDRL LNKLTTRQV EKNVQETRRR SQSSQFKAHA QREWENKTGL
 251 DFNHFIGGDI NKKGAVTGGH SLTRGDVRVI QQTSAADKHG VLSSDSGN*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

40
 1 TTGGGCATT CCCGAAAAT ATCCCTTATT CTGTCCATAC TGGCAGTGTG
 51 CCTGCCGATG CATGCACACG CCTCAGATTG GGCAACGAT CCCTTTATCC
 101 GgCaggttcT CGaccGTCAG CATTTCGaac ccgacggGAa ATACCAcCTA
 151 TTcggCaGCA GGGGGGAGCT TgcnagcGC aacggccATa tcggattggG
 201 aaacaTAcAa Agccatcagt tGggccacct gatgattcaa caggcggccg
 251 ttgaaggaaA TAtcgGctac attgtccgct tttccgatca cgggcacaaa
 301 ttccattcgc ccttcGAcaa ccaTGCCTCA CATTCCGATT CTGACGAAGC
 351 CGGTAGTCCC GTTGACGGAT TCAGCCTTTA CCGCATCCAT TGGGACGGAT
 401 ACGAACACCA TCCGCGCGAC GGCTATGACG GGCCACAGGG CGGCGGCTAT
 451 CCCGCTCCCA AAGGCGCGAG GGATATATAC AGCTACGACA TAAAAGGCGT
 501 TGCCCAAAAT ATCGCCTCA ACCTGACCGA CAACCGCAGC ACCGACAAAC
 551 GGCTTGCCGA CGTTTCCAC AATGCCGGCG CTATGCTGAC GCAAGGAGTA
 601 GGGCAGCGAT TCAAACGCGC CACCCGATAC AGCCCGAGC TGGACAGATC
 651 GGGCAATGcC gccGAAGCCT TCAACGGCAC TGCAGATATC GTCAAAAACA
 701 TCATCGGCGC GGCAGGAGAA ATTGTGCGCG CAGGCGATGC CGTGCagGGT
 751 ATAAGCGAAG GCTCAAACAT TGCTGTCAAT CACGGCTTGG GTCTGCTTTC
 801 CACCGAAAAC AAGATGGCGC GCATCAACGA TTTGGCAGAT ATGGCGCAAC
 851 TCAAAGACTA TGCCGCAAGC GCCATCCGCG ATTGGGCAGT CCAAAACCCC
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA TGGCAGCCAT
 951 CCCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGGCGGCA
 1001 TCACGGCACA TCCTGTCAAG CGGTGCGAGA TGGGCGCGAT CGCATTGCCG
 1051 AAAGGGAAAT CCGCGTCAG CGACAATTTC GCCGATGCGG CATACGCCAA
 1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC

-275-

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
 1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
 5 1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
 1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
 1451 CTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAAATAAA
 1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
 10 1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
 1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
 1651 ATTAATAAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
 1701 AGTGATGACC AAGCACACCA TGTTCCTCAA AGATTGGGAT GAGGCTAGAA
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
 1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAATAATAG AAGGATTTAC
 15 1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKYHL
 51 FGSRGELAXR NGHIGLGNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGHK
 101 FHSFPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
 20 151 PAPKGARDIY SYDIKGVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIIGAAGE IVGAGDAVQV
 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQIEAVS NIFMAAIPK GIGAVRGKYG LGGITAHVPK RSQMGAIALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 25 401 KNVKLADQRH PKTGVPPFDGK GPNFEKHVK YDTKLDIQEL SGGGIPKAKP
 451 VFDKPRWEV DRKLNKLTR EQVEKNVQET RRRSQSSQFK AHAQREWENK
 501 TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQOTSAPD KHGVYQATVE
 551 IKKPDGSWEV KTKKGGKVM THTMFPKDW EARRAEVTS AWESRIMLKD
 601 NKWQTSKSG IKIEGFTEPN RTAYPIYE*

30 ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40
orf46-1.pep		AVCLPMHAHASXLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER			
orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR				
35		10	20	30	40
		50	60	70	80
orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFSHDGHEVHSPFDNHAHSDSDEAGSP				
40	NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSHDGHHKFSFPDNHAHSDSDEAGSP				
		70	80	90	100
		110	120	130	140
orf46-1.pep	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTNRS				
45	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRLNLTNRS				
		130	140	150	160
		170	180	190	200
orf46-1.pep	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE				
50	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDRSGNAAEAFNGTADIVKNIIGAAGE				
		190	200	210	220
		230	240	250	260
55		270	280	290	300
orf46-1.pep	I				
orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP				
60		250	260	270	280

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of

N. meningitidis:

10 20 30 40 50 60

-276-

5	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDROHFEPDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFEPDGKYHLFGSRGELAXR
10	orf46a.pep	SGHIGLGNIQSHQLGNLFIQQAIAKGNIGYIVRFSHDGHEVHSPFDNHASHSDSDEAGSP
	orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSHDGHEVHSPFDNHASHSDSDEAGSP
15	orf46a.pep	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVQNIIRNLNDNRS
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVQNIIRNLNDNRS
20	orf46a.pep	TGQRLVDRFHNTGSMITQGVGDGFKRATRYSPELDRLSGNAEAFNGTADIVKNIIGAAGE
	orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDRLSGNAEAFNGTADIVKNIIGAAGE
25	orf46a.pep	IVGAGDAVQGISSEGSNIAMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
	orf46ng-1	IVGAGDAVQGISSEGSNIAMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
30	orf46a.pep	NAAQGIEAVSNIFTAVIPVKIGIAVRGKYGLGGITAHVPVKRSQMGELALPKGKSAVSDNF
	orf46ng-1	NAAQGIEAVSNIFMAAIPKIGIAVRGKYGLGGITAHVPVKRSQMGALALPKGKSAVSDNF
35	orf46a.pep	ADAAAYAKYPSPYHSRNIIRSNLEQRYGKENITSSTVPPSNGKNVKLANRHPKTKVPFDGK
	orf46ng-1	ADAAAYAKYPSPYHSRNIIRSNLEQRYGKENITSSTVPPSNGKNVKLADQRHPKTKVPFDGK
40	orf46a.pep	GFPNFEKDVKYDTRINTAVPQVN----PIDEPVFN--PKGSVGSASHWSITARIQYAKLP
	orf46ng-1	GFPNFEKHVKYDTKLD--IQELSGGGIPKAKPVFDAKPRWEVDKRLN-KLTTRQVEKQNV
45	orf46a.pep	RQGRIRYIPPKNYSPAPLPKGPNNGYLDKFGNEWTKGPSRTKGQFEWDVQLSKTGREQ
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDVRVLIQOTS

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCC	GCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CAT	GCACACG	CCTCAGATT	GGCAAACGAT	TCTTTTATCC
60	101	GGCAGGTTCT	CGAC	CGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGG	GGAAT	TGCCGAGCGC	AGCGGTCATA	TCCGATTGGG
65	201	AAACATACAA	AGCC	ATCAGT	TGGGCAACCT	GTTTATCCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATC	GGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCAGCAA
70	301	GTCCATTCCC	CCTT	CGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTG	ACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
75	401	ACGAACACCA	TCCG	CCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCCTCCCCA	AAGG	CGCGAG	GGATATATAC	AGCTACGACA	TAAAAGGCGT
80	501	TGCCCAAAAT	ATCC	GCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAAC
	551	GGCTTGTCGA	CCGT	TTCCAC	AATACCGGTA	GTATGCTGAC	GCAAGGAGTA
85	601	GGCGACGGAT	TCAA	ACGCGC	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCG	AAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
90	701	TCATCGGCGC	GGCA	GGAGAA	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTC	AAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
95	801	CACCGAAAC	AAGAT	TGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC

851 TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CAAAACCCC
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAAG AATATCTTTA CGGCAGTCAT
 951 CCCCGTCAAA GGGATTGGAG CTGTTCCGGG AAAATACGGC TTGGGCGGCA
 1001 TCACGGCACA TCCTGTCAAG CGGTCGCAGA TGGGCGAGAT CGCATTGCCG
 5 1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATACGCCAA
 1101 ATACCCGTCC CCTTACCATT CCCGAAATAT CCGTTCAAAC TTGGAGCAGC
 1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA
 1201 AAGAATGTGA AACTGGCAAA CAAACGCCAC CCGAAGACCA AAGTGCCGTT
 10 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA
 1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT
 1351 AATCCTAAG GTTCTGTGCG ATCGGCTCAT TCTTGGTCTA TAACTGCCAG
 1401 AATTCAATAC GCAAAATTAC CAAGGCAAGG TAGAATCAGA TATATCCAC
 1451 CTAAAAATTA CTCTCCTTCA GCACCGCTAC CAAAAGGACC TAATAATGGA
 1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAATAA
 15 1551 AGGTCAAGAA TTGAATGGG ATGTTCAATT GTCTAAAAA GGAAGAGAGC
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
 1651 GGAAAGATTA CACACAAATG A

This corresponds to the amino acid sequence <SEQ ID 466>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIRQVLDRO HFEPDGKYHL
 20 51 FGSRGELAER SGHIGLGNIQ SHQLGNLFIQ QAAIKGNIGY IVRFSDHGHE
 101 VHSPPDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
 151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLVDRFH NTGSMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAENGTAADI VKNIIGAAGE IVGAGDAVQG
 25 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQGIEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGEIALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVE QVNPIDEPVF
 451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPKNYSPS APLPKGPNNG
 501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
 30 551 GKITHK*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

1 ATGAATATTC ACACCCTGCT CTCCAACAA TGGACGCTGC CGCCATTCTCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 40 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTTTCGT
 201 CAAAATTGCC GCGTATTGG CGTTTGGCT GCGGTTTTTG TTTGACGGGC
 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CTTTCATCCT GACCGCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTG...

45 This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWVLAL LTATARPIVN
 51 LDYLPALLI ALPWRFKIA GVLAFWLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

50 1 ATGAATATTC ACACCCTGCT CTCCAACAA TGGACGCTGC CGCCATTCTCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTTTCGT
 201 CAAAATTGCC GCGTATTGG CGTTTGGCT GCGGTTTTTG TTTGACGGGC
 55 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CTTTCATCCT GACCGCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTGTG TTGCAGAAAG

-278-

5
10
15
20

```

401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTGC
501 GATGGCCAAT ATCTTCGGCG CAAACAACCT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTCGCGCG GTTCGCACTG CGCCGCGCGC CGGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CCGCGCGGGC AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGGTG GCGACGCGAG CTGTTCCGGC AAGTGTGCGC
1101 ATTTTTCAAA AAACACGACA AGGGACTGTT TTAAGTGGAT ACGCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATTT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTCAGCCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
1301 TGAAGGCAC GGAAGTCATC ATCGTCGCGC ACCATCCGCC GCCCGTCGCG
1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGCCTGGCT
1401 GAACCTCAA ATCAAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLSLLILLA PNAVFWVLAL LTATARPIVN
51 LDYLPALLI ALPWRVFKIA GVLAFLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADEFITAGL
201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNAFPAKL
251 LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFA
301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKTC
351 ATEGGVCDSE LFEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHRLKC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVVG
451 NLNETFRYLK QGHVAWLNEK IK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N.*

meningitidis:

40
45
50

```

              10      20      30      40      50      60
orf48.pep    MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPALLI
              |||||
orf48a       MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWVLALLTATARPIVNLXYLPALLI
              10      20      30      40      50      60

              70      80      90      100     110     119
orf48.pep    ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
              |||||
orf48a       ALPWRXVKIXGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFIXTAPALYQIMTGLL
              70      80      90      100     110     120

orf48a       LLYMLAMPFVLQKAAKTDFRHIAACAIVVVAAGYFTGHLSXYDRGRMANIFGANNFYA
              130     140     150     160     170     180

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55
60

```

1 ATGAATATTC ACACCTGCT CTCCAACAA TGGACGCTGC CGCCATTCCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCCG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
151 TLTGGANTACC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GCGGTNTCGT
201 CAAAATTGNC GCGGTATTGG CGTNTTGGCT GCGGTTTTTG TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGATCTCAT CGGCGCCATC
301 AACCTCGTCC CTTTCATCNT GACCGCCCCC GCCCTTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCCTGTGC CGCCGTTGTG
451 GTGCGAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGCG

```

501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTATTACGCC AAAAGTCAGG
 551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
 601 TCTGATCCCG TCTTCCTCCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
 651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
 751 CTGGCGCAAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCTT
 801 CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
 851 GTTTCGCGCG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
 901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
 951 CGGCGCGGCG AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
 1001 GCTTTCGAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTTCGGCG AAGTGTCCGC
 1101 ANTTTTTCAA AAACACGACA AGGGAAGTGT TTAGTGGATG ACGCTGACCA
 1151 GCCACGCCGA CTATCCCGAA TCNGACATTT TCAACCACAG GCTCAAATGC
 1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTCAGCCT
 1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCCGAAA
 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
 1351 AACCTCAATG AACCTTCCG CTACCTCAAA CAGGGGCACG TCGNCTGGCT
 1401 GAACCTTCAA ATCAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPLPKR LLLSLILLX PNAVFWLAL LTATARPIVN
 51 LXLYPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFIFTAP ALYQIMTGLL LLYMLAMPFV LQKAAAKTDF RHIAACAAYV
 151 VAAGYFTGHL SXYDRGRMAN IFGANNFYFA KSQAMLYTVS QNADFITAGL
 201 VDPVFLPLGN QQRAATHLNE PKSQKILFIV AESWGLPANP ELQNATFAKL
 251 LAQKXRFSVW ESGSFPIGA TIEGEMREL CAYGGLRGFAL RRAPDEKFA
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKKT
 351 AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYPE SDIFNHLK
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVP
 451 NLNETFRYLK QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
orf48a.pep		MNIHTLLSKQ	WTLPPLPKR	LLLSLILLX	PNAVFWLAL	LTATARPIVN	LXLYPAALLI
35	orf48-1	MNIHTLLSKQ	WTLPPLPKR	LLLSLILLX	PNAVFWLAL	LTATARPIVN	LDYLPALLI
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf48a.pep	ALPWRXVKIX	GVLAXWLAVL	FDGLMMVIQL	FPFMDLIGAI	NLVPFIFTAP	ALYQIMTGLL
	orf48-1	ALPWRXVKIX	GVLAXWLAVL	FDGLMMVIQL	FPFMDLIGAI	NLVPFILTAP	ALYQIMTGLL
		70	80	90	100	110	120
45	orf48a.pep	LLYMLAMPFV	LQKAAAKTDF	RHIAACAAYV	VAAGYFTGHL	SXYDRGRMAN	IFGANNFYFA
	orf48-1	LLYMLAMPFV	LQKAAAKTDF	RHIAACAAYV	VAAGYFTGHL	SXYDRGRMAN	IFGANNFYFA
		130	140	150	160	170	180
50	orf48a.pep	KSQAMLYTVS	QNADFITAGL	VDPVFLPLGN	QQRAATHLNE	PKSQKILFIV	AESWGLPANP
	orf48-1	KSQAMLYTVS	QNADFITAGL	VDPVFLPLGN	QQRAATHLNE	PKSQKILFIV	AESWGLPANP
		190	200	210	220	230	240
55	orf48a.pep	ELQNATFAKL	LAQKXRFSVW	ESGSFPPIGA	TIEGEMREL	CAYGGLRGFAL	RRAPDEKFA
	orf48-1	ELQNATFAKL	LAQKXRFSVW	ESGSFPPIGA	TIEGEMREL	CAYGGLRGFAL	RRAPDEKFA
60		250	260	270	280	290	300
	orf48a.pep	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE
65	orf48-1	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQEIKT	AENLIGKKT	CAIFGGVCDSE
		310	320	330	340	350	360

-280-

		370	380	390	400	410	420
orf48a.pep		LFEVSA	XFKKHKGLFYWMTLTSHADY	PESDI	FNHRLKCTEYGLPAETDXCRN	FSLHTQ	
orf48-1		LFEVSA	XFKKHKGLFYWMTLTSHADY	PESDI	FNHRLKCTEYGLPAETDL	CRNFSLHTQ	
5		370	380	390	400	410	420
		430	440	450	460	470	
orf48a.pep		FFDQLADLIQRPEMKGTEVIIVGDHPPPVG	NLNETFRYLKQGHVXWLNFKIKX				
orf48-1		FFDQLADLIQRPEMKGTEVIIVGDHPPPVG	NLNETFRYLKQGHVXWLNFKIKX				
10		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPAAALI	60
		: : : : : : : : : :	
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPIVNLDYLPAAALI	60
20	orf48.pep	ALPWRFVKIAGVLAFWLAFLDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
		: : : : : : : : :	
	orf48ng	ALPWRFVKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	PNAVFWVLAL	LTATARPIVN
	51	LDYLPAAALI	ALPWRFVKIA	GVLAFWPAVL	FDGLMMVIQL	FPFMDLIGAI
	101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LQKAAVKTD	RHIAVCAAVV
	151	AAARYFTGPF	ELLRTGGRWQ	YVQHRRLLS	GSRASFRRRQ	KADVLRRRLGN
	201	PYASMGNGG

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCTCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCGG
	101	TGTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCGCGCC	GATTGTCAAT
35	151	TTGGACTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
	201	CAAAATTGCC	GGCGTATTGG	CGTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAACATC	TTCCCTTTTA	TGGACCTCAT	CGCGGCCATC
	301	AACCTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCTTATC	AGATAATGAC
	351	CGGCTGTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTGTG	TTGCAAAAAG
40	401	CCGCCGTCAA	AACCGACTTC	CGACACATTC	CCGTCTGTGC	CGCCGTTGTG
	451	GCGGCAGCCG	GCTATTTTAC	CGGCCATTTC	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGGCG	CAACAACAT	CTATTACGCC	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAAATGCC	ACTTTATTAC	CGCCGgctG
	601	GTCGACCCCG	TCTTCTCTCC	CTTGGGCAAT	CAGCAGCGTG	CCGCCACGCG
45	651	GCTGAGTGAG	CCGAAATCTC	AAAAAATCCT	CTTTATCGTC	GCCGAATCTT
	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCCTT
	801	CATCGGCGCG	ACGGTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTTCGCGG	GTTTCGCACTG	CGCCGCGCGC	CCGACGAAAA	ATTTGCCCGC
50	901	TGCCTCCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
	951	CGGCGCGGGT	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGTGTG	CGACAGCGAG	CTGTTTCGGC	AAGTGTGCGC
	1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACCGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTCAGCCT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	GCCGTCGGC
	1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGACACG	TCGCCTGGCT
	1401	GCACTTCAAA	ATCAATAA			

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

```

      1 MNIHALLSEQ WTLPPFLPKR LLLSLLILLA PNAVFVWLAL LTATARPIVN
    51 LDYLPALLI ALPWRVFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
   101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKA AVKTD F RHI AVCAAVV
   151 AAAGYFTGHL SYDRGRMAN IFGANNFYIA KSQAMLYTVS QNADFITAGL
    5 201 VDPVFLPLGN QORAATRLSE PKSQKILFIV AESWGLPGNP ELQNATFAKL
   251 LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFAR
   301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQIKT AENLIGKKT
   351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
   401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPVVG
  10 451 NLNETFRYLK QGHVAVLHFK IK*

```

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

```

      10      20      30      40      50      60
  orf48-1.pep MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARPIVNLDYLPALLI
  15 orf48ng-1 MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARPIVNLDYLPALLI
      10      20      30      40      50      60

      70      80      90     100     110     120
  orf48-1.pep ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL
  20 orf48ng-1 ALPWRVFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL
      70      80      90     100     110     120

      130     140     150     160     170     180
  orf48-1.pep LLYMLAMPFVLQKAAAKTDFRHI AVCAAVVAAAGYFTGHLSYYDRGRMANIFGANNFYIA
  25 orf48ng-1 LLYMLAMPFVLQKAAVKTDFRHI AVCAAVVAAAGYFTGHLSYYDRGRMANIFGANNFYIA
      130     140     150     160     170     180

      190     200     210     220     230     240
  orf48-1.pep KSQAMLYTVSQNADFITAGLVDPVFLPLGNQORAATHLNEPKSQKILFIVAESWGLPANP
  30 orf48ng-1 KSQAMLYTVSQNADFITAGLVDPVFLPLGNQORAATRLSEPKSQKILFIVAESWGLPGNP
      190     200     210     220     230     240

      250     260     270     280     290     300
  orf48-1.pep ELQNATFAKLLAQKDRFSVWESGSPFFIGATVEGEMRELCA YGGLRGFALRRAPDEKFAR
  35 orf48ng-1 ELQNATFAKLLAQKDRFSVWESGSPFFIGATVEGEMRELCA YGGLRGFALRRAPDEKFAR
      250     260     270     280     290     300

      310     320     330     340     350     360
  orf48-1.pep CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEI KTAENLIGKKTCAIFGGVCDSE
  45 orf48ng-1 CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQKIKTAENLIGKKTCAIFGGVCDSE
      310     320     330     340     350     360

      370     380     390     400     410     420
  orf48-1.pep LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDI FNHRLKCTEYGLPAETDL CRNFSLHTQ
  50 orf48ng-1 LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDI FNHRLKCTEYGLPAETDL CRNFSLHTQ
      370     380     390     400     410     420

      430     440     450     460     470
  orf48-1.pep FFDQLADLIQRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAVLNFKIKX
  55 orf48ng-1 FFDQLADLIRPEMKGTEVIIVGDHPPPVGNLNETFRYLKQGHVAVLHFKIKX
      430     440     450     460     470

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined)

60 and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCCAAAA TCATCATCGT
      51  TACTTTGAGT ATCGCCACGC TTGCCGCGCG CGGCATCGCT ATGTCGCGCG
5     101  GTATGCGAGT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
      151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
      201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
      251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
      301  AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGGCGC.G TAGCGCCGAA
10    351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAAATAT AACGGGCAAT
      401  TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMSD FIEPTPWTLA
      51  GLGFLIALMG WMPAIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15    101  SAVLALVFLA LGXVAPNGNG XTVQMAGGKY NGQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1  ATGTCCGAAC AACATATTTT GACTTGGAAA AGTAAATCA ACGCATTGGG
      51  TCCGGGGGATC ATGATGGCTT CGCGCGCGGT CGCGGGTTCG CACCTGATTG
20    101  CCTCGACGCA GGCGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
      151  ATCCTGACCA ACCTCTTCAA ATACCGGTTT TTCCGCTTCA GCGCGCATTA
      201  CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
      251  GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
      301  AACGCGGGCG CGGTCGCCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
      351  TCCCTCGCTG ATGTTTGTATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
25    401  CTGCGCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
      451  TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
      501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
      551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
      601  ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
30    651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTC
      701  ACGTCGGTTA TATCGCCAGT GCGGTTTGG CTTTGGTTT CCTTGCACTG
      751  GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
      801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTACC ATCGGCGGCT
      851  GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
35    901  CAGGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
      951  CCGTCTGCGC GGAAAAGACA AAACGGGCAA CGCCGAATTC TTTGCCTGGA
1001  ATATTTGGGT GGCGGGCAGC GGTTTGGCGG TGATTTTCTG GTTTGACGGC
1051  GTAATGCGGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101  CCCTGTGTTT GCCTGGCTGA ATTACCGTTT GGTTAAAGGT GATGAAAAAC
40    1151  ACAAACTCAC ATCAGGTATG AATGCCCTTG CATTTGGCAG CTTGATTAT
1201  CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251  ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
45    51  ILTNLFKYFF FRFSAHYTLD TGKSLIEGYA EKSRYVLWVF LILCILSATI
      101  NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
      151  SKIIIVTLSI ATLAAAGIAM SRGMQMSDF IEPTPWTLAG LGFLIALMGW
      201  MPAPIEISAI NSLWVTEKQR INPSEYRDI FEFNVGYIAS AVLALVFLAL
      251  GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
50    301  TTTVVDGYAR AIAEPVRLR GKDKTGNAEF FAWNIIWVAGS GLAVIFWFDG
      351  VMANLLKFAM IAAVFSAPVF AWLNYRLVKG DEKHKLTSGM NALALAGLIY
      401  LTGFTVLFLL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N.meningitidis*:

-283-

					10	20	30
	orf53.pep				VSGRYRALDRVSKIIIVTLSIATLAAAGIA		
5	orf53a	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVS	SGRYRALDRVSKIIIVTLSIATLAAAGIA				
		110	120	130	140	150	160
	orf53.pep		40	50	60	70	80
	orf53a	MSRGMQMQSDFIEPTPWTLAGLGLIALMGWMPAP	IEISAINSLWVTEKQRINPSEYRDG				
10		170	180	190	200	210	220
	orf53.pep		100	110	120	130	139
	orf53a	IFEFNVGYIASAVLALVFLALGXVAPNGXTVQ	MAGGKYNQLINMYA				
15		230	240	250	260	270	280
	orf53a	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQ	MAGGKYIGQLINMYAVTIGGWSRPLV				
20	orf53a	AFIAFACMYGTTITVVDGYARAIAPVRLRLRGK	DKTGNAEFFAWNIVVAGSGLAVIFWFD				
		290	300	310	320	330	340

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

	1	ATG	TCCGAAC	AACATAT	TTTC	GACTTGG	AAA	AGTAAAATCA	ACGCATTGGG
	51	ACC	GGGGATT	ATGATGG	CCT	CGGCGGCGGT	CGGCGGTT	CGC	CACCTGATTG
25	101	CCT	CGACGCA	GGCGGG	CGCG	CTTTACGGCT	GGCAGATCGC	GCTCATCATC	
	151	ATC	CTGACCA	ACCTCTT	CAA	ATACCCGTTT	TTCCGCTTCA	GCGCGCATT	
	201	CAC	GTGGAC	ACGGGCA	AAGA	GCCTGATTGA	AGGTTATGCC	GAGAAAAGCC	
	251	GCG	TTTATTT	GTGGGT	TATTC	CTGATTTGT	GCATCCTCTC	CGCCACGATT	
	301	AAC	CGGGCG	CGTCCG	CAT	TGTAACCGCC	GCCATCGTCA	AAATGGCGAT	
	351	TCC	CTCGCTG	ATGTTG	ATG	CCGGCACGGT	TGCCGCCTTG	ATTATGGCAT	
30	401	CCT	GCCTGAT	TATTTG	GTG	AGCGGACGTT	ACCGCGCTTT	GGATCGCGTT	
	451	TCC	AAAATCA	TCATCGT	TAC	TTTGAGTATC	GCCACGCTTG	CCGCCGCGG	
	501	CAT	CGCTATG	TCGCGC	GTA	TGCAGATGCA	GTCCGATTTT	ATCGAGCCGA	
	551	CAC	CGTGGAC	GCTTGC	CGGT	TTGGGCTTCC	TGATCGCGCT	GATGGGCTGG	
	601	ATG	CCCGCGC	CGATTGA	AAAT	TTCCGCCATC	AATTTCTTGT	GGGTAACCGA	
35	651	AAA	ACAACGC	ATCAAT	CCTT	CCGAATACCG	CGACGGGATT	TTTGATTTC	
	701	ACG	TCGGTTA	TATCGC	CGAGT	GCGGTTTGG	CTTTGTTTTT	CCTTGCCTG	
	751	GGC	CGGTTTG	TGCAAT	ACGG	CAACGGCGAA	GCAGTGCAGA	TGGCGGGCGG	
	801	CAA	TATATC	GGGCA	ATTGA	TCAATATGTA	CGCCGTTACC	ATCGGCGGCT	
	851	GGT	CGCGCCC	GCTGGT	GGCG	TTTATCGCGT	TTGCCTGTAT	GTACGGCACG	
40	901	ACG	ATTACCG	TTGTGG	ACGG	CTATGCCCGT	GCCATTGCCG	AACCCGTGCG	
	951	CCT	GTGCGC	GGAAA	GACA	AAACGGGCAA	CGCCGAATTC	TTTGCCTGGA	
	1001	AT	ATTGGGT	GGCGGC	CAGC	GGTTTGGCGG	TGATTTTCTG	GTTTGACGGC	
	1051	GTA	ATGGCGA	ATCTG	CTCAA	ATTTGCGATG	ATTGCGGCTT	TTGTGTCCGC	
	1101	CC	CTGTGTTT	GCCTGG	CTGA	ATTACCGTTT	GGTCAAAGGT	GATGAAAAAC	
45	1151	ACA	AACTCAC	ATCAG	GATG	AATGCCCTTG	CATTGGCAGG	CTTGATTAT	
	1201	CT	GACCGGT	TTACCG	TTTT	GTTCTTATTG	AATTTGGCGG	GAATGTTCAA	
	1251	ATGA							

This encodes a protein having amino acid sequence <SEQ ID 482>:

	1	MSE	QHISTWK	SKINALGPGI	MMASAAV	GGSHLIA	STQAGALY	GWQIALII	
50	51	IL	TNLFKYPF	FRSAHYTLD	TGKSLIEGYA	EKSRVYLWVF	LILCILSATI		
	101	NAG	AVAVITA	AIVKMAIPSL	MFDAGTVAAL	IMASCLIIIV	SGRYRALDRV		
	151	SK	IIIVTLSI	ATLAAAGIAM	SRGMQMQSDF	IEPTPWTLAG	LGFLIALMGW		
	201	MP	APIEISAI	NSLWVTEKQR	INPSEYRDGI	FDNFVGYIAS	AVLALVFLAL		
	251	GA	FVQYNGE	AVQMAGGKYI	GQLINMYAVT	IGGWSRPLVA	FIAFACMYGT		
55	301	TIT	VVDGYAR	AIAEPVRLLR	GKDKTGNAEF	FAWNIWVAGS	GLAVIFWFDG		
	351	VM	ANLLKFAM	IAAFVSAPVF	AWLNRYRLVKG	DEKHKLTSGM	NALALAGLIY		
	401	LT	GETVLFL	LL	NLAGMEF*				

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

		10	20	30	40	50	60
60	orf53a.pep	MSEQHISTWKS	SKINALGPGIMMASAAVGGSHLIA	STQAGALYGWQIALII	ILTNLFKYPF		
	orf53-1	MSEQHISTWKS	SKINALGPGIMMASAAVGGSHLIA	STQAGALYGWQIALII	ILTNLFKYPF		
		10	20	30	40	50	60
65		70	80	90	100	110	120

-284-

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAAIVKMAIPSL	
5		70 80 90 100 110 120	
	orf53a.pep	130 140 150 160 170 180	
	orf53-1	MFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTSLIATLAAAGIAMS RGMQM QSDF	
10		130 140 150 160 170 180	
	orf53a.pep	190 200 210 220 230 240	
	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS	
15		190 200 210 220 230 240	
	orf53a.pep	250 260 270 280 290 300	
	orf53-1	AVLALVFLALGAFVQYNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT	
20		250 260 270 280 290 300	
	orf53a.pep	310 320 330 340 350 360	
	orf53-1	TITVVDGYARAIAPVRLLRGKDKTGNAEFFAWNIWVAGSGLAVIFWFDGVMANLLKFAM	
25		310 320 330 340 350 360	
	orf53a.pep	370 380 390 400 410	
	orf53-1	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMFKX	
30		370 380 390 400 410	
	orf53a.pep	IAAFVSAPVFAWLNYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVFLNLAGMFKX	
35		370 380 390 400 410	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTSLIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIILVSGRYRALDRVSKIIIVTSLIATLAAAGIA	91
	orf53.pep	MSRGMQM QSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
45	orf53ng	MSRGMQM QPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGXGTVQMAGGKYNGQLINMYA	139
50	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYNGEAVQMGGGKYIGQLINMYAVTIGGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMOP
	101	DFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGGSRPL VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARP
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20

25

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

30

35

40

45

50

55

60

65

	60	70	80	90	100	110
orf53-1.pep	ILTNLFKYPPFFRFS	AHYTLDTGKSL	IEGYAEKSRVYL	WVFLILCILSAT	INAGAVAIVTA	
orf53ng-1			:			
				KKSCVYLWVFLIL	CIASATINAGAVAIVTA	
				10	20	30
orf53-1.pep	120	130	140	150	160	170
orf53ng-1	AIVKMAIPSLMFDAGTVAALIMASCLII	LVSGRYRALDRVSKIIIVT	LSIATLAAAGIAM			
orf53-1.pep	120	130	140	150	160	170
orf53ng-1	AIVKMAIPSLMFDAGTVAALIMASCLII	LVSGRYRALDRVSKIIIVT	LSIATLAAAGIAM			
	40	50	60	70	80	90
orf53-1.pep	180	190	200	210	220	230
orf53ng-1	SRGMQMQSDFIEPTPWTL	LAGLGFLIALMGWMPAP	IEISAINSLWVTEKQ	RINPSEYRDGI		
orf53-1.pep	180	190	200	210	220	230
orf53ng-1	SRGMQMQPDFIEPTPWTL	LAGLGFLIALMGWMPAP	IEISAINSLWVTEKQ	RINPSEYRDGI		
	100	110	120	130	140	150
orf53-1.pep	240	250	260	270	280	290
orf53ng-1	FDFNVGYIASAVLALVFL	ALGAFVQYNGEAVQMAGG	KYIGQLINMYAVTIGGWS	RPLVA		
orf53-1.pep	240	250	260	270	280	290
orf53ng-1	FDFNVGYIASAVLALVFL	ALGAFVQYNGEAVQMAGG	KYIGQLINMYAVTIGGWS	RPLVA		
	160	170	180	190	200	210
orf53-1.pep	300	310	320	330	340	350
orf53ng-1	FIAFACMYGTTITVVDGYARAIAEP	VRLLRGRDKTGNAELFAWNIWVAGSGLAVIFW	FDG			
orf53-1.pep	300	310	320	330	340	350
orf53ng-1	FIAFACMYGTTITVVDGYARAIAEP	VRLLRGRDKTGNAELFAWNIWVAGSGLAVIFW	FDG			
	220	230	240	250	260	270
orf53-1.pep	360	370	380	390	400	410
orf53ng-1	VMANLLKFAMIAAFVSAPVFAWL	NYRLVKGDEKHKLTSGMNALALAGLIYLTGFTVLF	ELL			
orf53-1.pep	360	370	380	390	400	410
orf53ng-1	VMANLLKFAMIAAFVSAPVFAWL	NYRLVKGDKRHRLTAGMNALAI	VGLLYLAGFAVLF	ELL		
	280	290	300	310	320	330
orf53-1.pep	NLAGMFKX					
orf53ng-1	NLTGLLAX					

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
51  TGGCCTTGCC GGCTTGTTTT TTGTCCGCGC ACAATCCGAA CGCGAGTGGA
101 TGGCGGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGCGGAG
151 CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCTGAT
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGTG TTGGCTGTTT GTCGGTGTG
251 TCCGTTTCTG CCGAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
301 GTTCCGCCT..

```

This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QEKKGKQAE
51  LPEIKDGMPD FPELALMLFH AVKTAVYWLF VGVVRFERNY LAHESEPDPR
101  VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTGTGA TAGTTTGTAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTGTGC CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTTGGC TGTTTGTGCG TGTCGTCCGT TTCTGCCGAA
25  251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
30  501 AATTTGCCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCGGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
35  751 TCTGCCGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACCGTC
851 ATGCGAGGCA GGGGAAAGGG CAGGCGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGCGAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACC GAACAA GGCAACGGTT TCTGCGGAGG
40  1001 CGCGAATTTT TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAAACGG
1051 GATGTCGAAA TGCCGTCTGA AACC GAAAT GTTTTCACGG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAGTACC GATATCCATA
1151 TTGAAGAACG TGCCGCGCCC GATGCTTGGG TGCTCGAACC ACCCGAAGTG
45  1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACC GTACCT ATGAACGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCGGAG ACCGACCATC TTGCCGATGA TGTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAACCC GAAGCGTTCTG
1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
50  1501 CCGTCTTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAATGTTG
1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
70  1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT AATTACGCGT TATGAAATCG
1751 AATCCGATGT CGGCGTGC GC AATTCG TTCTGAATCT GGAAAAAGAT
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTG AAACCATCCC
1851 CGGCAAAACC TGCATGGGTT TGGAACCTCC GAACCGGAAA CGCCAAATGA

```

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
 2001 CGACTTGGGA AAAGCACCGC ATTTGTTGGT TGCCGGCACG ACCGGTTCGG
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTG TGTCTATGCT TTTCAAAGCC
 5 2101 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCCGTGC GTTACCGATA
 2201 TGAAGCTGGC GGCAAACGCG CTGAACCTGG GTGTTAACGA AATGGAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CGGGCTTCAA
 10 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
 2351 TCAGCCTCAC GCCCAGCATG CCCGAACCTT TGGAAAAACT GCCGTTTATC
 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
 2501 TCCATTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
 15 2601 AATCGACAGC CGCACGATTG TCGACCAAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGGCGA TATGCTGTTC CTGCTGCCGG GTACTGCCTA TCCGCAGCGC
 2701 GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAGCGA AGAGCTGCCG GGCATCGGGC GCAGCGGCGA CGACGAACCC
 20 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
 2901 CAGCATTTTC GCGCTACAGC GCGCCTTGCG TATCGGCTAC AACCGCGCCG
 2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFIVILVILV LLAGLFFV RAQSEREWMR EVSAWQEKKG EKQAELEPEIK
 51 DGMPOFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
 101 ANRADVPTAS DGYSDSNGT EEAETEEAEA AEEEEADTED IATAVIDNRR
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQGKG QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR
 351 DVEMPSETEN VFTETVSSVG YGGPVYDETA DIHIEEPAAP DAWVVEPPEV
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPSPGF EQVQRSRIAE TDHLADDVLN
 451 GGWQEETAAI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CPFENVPSER
 501 PSCRVSDETA DEGAFFSEET GAVSEHLPTT DLLLPLFNP EATQTEEELL
 551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGR GNSVLNLEKD
 601 LARSIGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 651 KLTLALGQDI TGQPVVTDLG KAPHLVAGT TSGSKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLAPV VTDMKLAANA LNWCVNEMEK
 40 751 RYRLSFMGV RNLAFNQKI AEAARGEKI GNPFSLTPDD PEPLKLPFI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LLPGTAYPQR
 901 VHGAFADEE VHRVVEYLKQ FGEPDYDDI LSGGSEELP GIGRSGDDET
 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
 45 1001 HNGNRTILVP LDNA*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

orf58.pep 10 20 30 40 50 60
 LRETAYVLDSFDYFVVALAGLFFVRAQSEREWMRVSAWQEKKGKQAELEPEIKDGMMPD
 orf58a 10 20 30 40 50
 MFIVILVILVLLLAGLFFVRAQSEREWMRVSAWQEKKGKQAELEPEIKDGMMPD
 55
 orf58.pep 70 80 90 100
 FPTELMLFHAVKTAVYWLFGVVVRFCRNYLAHESPEPDRPVPP
 orf58a 60 70 80 90 100 110
 FPTELMLFHAVKTAVYWLFGVVVRFCRNYLAHESPEPDRPVPPASANRADVPTASDGYS

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTG CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
5   GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
151 CAAAACGGCA GTGTATTTGGC TGTTTGTCGG TGTCTGTCGT TTCTGCCGAA
201 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
10  AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCCGCCG
401 ATCCCATTCC ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
451 AATTTTCGCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
501 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
551 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
601 TACCCCGATG GAAGGCTGTC AGATTATCGG TTTGGACGAC CCTGTGCTTC
15  AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
701 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCGTGA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCGCTC
851 ATGCCAGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
20  CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
901 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
951 CGCGGATTTC GCGCCTGATT CCGGAAAGTC GGACGGTTGT CCGGAAACGG
1001 GATGTGCGAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1051 GTCTGTGGGA TACGCGGNTC CGGTTTATGA TGAAACTGCC GATATCCATA
1101 TTGAAGAACG TGCCGCGCCC GATGCTTGGG TGGTTCGAAC ACCCGAAGTG
1151 CCGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
1201 GGAAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTG GAGCAGGTGC
1251 AACGCAGCCG CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTTGAAT
1301 GGAGGTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGGG
1351 TGTGGCAGAG CGGTCAGCG GGCAATATTT GTCGGAACG GAAGCGTTTCG
30  GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAAATGTGCC GTCTGAACGC
1451 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
1501 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1551 TGCCGCGCCT GTTCAATCCC GGGGCGACGC AAACCGAAGA AGANCTGTTG
35  GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1651 CAAGGTGTGC GATTCTTATT CCGGCCCCGT GATTACGCGT TATGAAATCG
1701 AACCCGATGT CCGCGTGC GC AATTTCCG TTCTAAATCT GGAAAAAGAN
1751 TTGGCGCGTT CGCTCGCGCT GGCTTCCATC CGCGTGTGCG AAACCATCCT
1801 CGGCAAAACC TGTATGGGTT TGGAACCTCC GAACCCGAAA CGCCAAATGA
40  TACGCTGAG CGAAATCTTC AATTGCCCCG AGTTTGCCGA ATCCAAATCC
1901 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
1951 CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCACG ACCGGTTCGG
2001 GCAAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
2051 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT
45  GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCATA
2151 TGAAGCTGGC GGCAACCGCG CTGAACCTGG GTGTTAACGA AATGGAAAAA
2201 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGTNTCAA
2251 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAAATC GGCAACCCGT
2301 TCAGCCTCAC GCCCGACAAT CCCGAACCTT TGGANAAATT GCCGTTTATC
2351 GTGGTCTGTT TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
50  AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2401 TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCAGGGT
2451 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2501 AATCGACAGC CGCAGATTTC TTGACCAAAT GGGTGCGGAA AACCTGCTCG
55  GGCAGGGCGA TATGCTGTTC CTGCCGCCG GTACGGCCTA TCCGACGCGC
2601 GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTCAATA
2651 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2701 GTATGTCCGA CGATTGCTG GGAATCAGCC GGAGCGGCGA CGGCGAAACC
2751 GATCCGATGT ACGACGAGGC CGTGTCTGTT GTTTTAAAA CGGCGAAAGC
2801 CAGCATTTCT GCGGTGACG GCGCATGCG TATCGGCTAT AATCGCGCCG
60  CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
2901 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA
3001

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

1  MFIVLVIL LLAGLEFFV RAQSEREWMR EVSAWQEKKG EKQAELEIK
65  DGMDFPELA LMLFHAVKTA VYWLFGVVVR FCRNYLAHES EPDRPVPPAS
51  ANRADVPTAS DGYSDSGNGT EEAETEAEAE AEEEAADTED IATAVIDNRR
101  IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
151  DAFEKNETAV PKVRVSDTPM EGLQIIIGLDD PVLQRTYSRM FDADKEAFSE
201

```

251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQGKG QAEAKSPDVS
 301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTEXVSSVG YGXPVYDETA DIHIEEPAAP WDAWVVEPPEV
 401 PKVPMXPAXDI PPPPVSEIY NRTYEPPAGF EQVQRSRIAE TDHLADDVLN
 5 451 GGWQEETA AI ANDGSEGVAE RSSGQYLSET EAFGHDSQAV CPFENVPSER
 501 PSRRAXDTEA DEGAFAQSEET GAVSEHLPTT DLLLPPLFNP GATQTEEXLL
 551 XNSITIEEKX AEFKVKVKV DSYS GPVITR YEIEPDVGVR GNSVLNLEKX
 601 LARSLGVASI RVVETILGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 10 651 KLTALGQDI TGQPVVTDLG KAPHLLVAGT TSGSKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLAANA LNWCVNEMEK
 751 RYRLMSFMGV RNLAGXNQKI AEAARGEKI GNPFSLTPDN PEPLXKL PFI
 801 VVVVDEFADL MMTAGKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LTKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR
 901 VHGAFAASDEE VHRVVEYLKQ FGE PDYVDDX LSGGMSDDL GISRSGDGET
 15 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
 1001 HNGNRTILVP XDNA*

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

20	orf58a.pep	10	20	30	40	50	60
	orf58-1	10	20	30	40	50	60
25	orf58a.pep	70	80	90	100	110	120
	orf58-1	70	80	90	100	110	120
30	orf58a.pep	130	140	150	160	170	180
	orf58-1	130	140	150	160	170	180
35	orf58a.pep	190	200	210	220	230	240
	orf58-1	190	200	210	220	230	240
40	orf58a.pep	250	260	270	280	290	300
	orf58-1	250	260	270	280	290	300
45	orf58a.pep	310	320	330	340	350	360
	orf58-1	310	320	330	340	350	360
50	orf58a.pep	370	380	390	400	410	420
	orf58-1	370	380	390	400	410	420
55	orf58a.pep	430	440	450	460	470	480
	orf58-1	430	440	450	460	470	480
60	orf58a.pep	490	500	510	520	530	540
	orf58-1	490	500	510	520	530	540

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKX					
5	orf58-1	EATQTEEELENSITIEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKD					
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIRVVETILGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
10	orf58-1	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
	orf58a.pep	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
15	orf58-1	TGQPVVTDLGKAPHLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720
	orf58a.pep	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLMSFMGVRNLAGXNQKIAEAAARGEKI					
20	orf58-1	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58a.pep	GNPFSLTPDNPEPLXKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
25	orf58-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58a.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITILDQMGAEENLLGQGDMLFLPGTAYPQR					
30	orf58-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRITILDQMGAEENLLGQGDMLFLPGTAYPQR					
		850	860	870	880	890	900
	orf58a.pep	VHGAFASDEEVHRVVEYLKQFGEPDYVDDXLSGGMSDDLGLISRSRGDGETDPMYDEAVSV					
35	orf58-1	VHGAFASDEEVHRVVEYLKQFGEPDYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58a.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
40	orf58-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

	orf58.pep	ALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPP	103
55	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	..SEPDRPVPPA	SANRADVPTA	SDGYSDSGNG	TEEAETAAE	AAAAEAADTE
60	51	DIATAVIDNR	RIPFDRSIAE	GLMQSESKTS	PVRPVFKEIT	LEEATRALSS
	101	AALRETKKRY	IDAFEKNGTA	VPKVRVSDTP	MEGLQIIGLD	DPVLQRTYSR
	151	MFDADKEAFS	ESADYGFEPY	FEKQHPSAFS	AVKAENARNA	PFRRHAGQEK
	201	GQAEAKSPDV	SQGQSVSDGT	AVRDARRRVS	VNLKEPNKAT	VSAEARISRL
	251	IPESRTVVGK	RDVEMPSETE	NVFTETVSSV	GYGGPVYDEA	ADIIHIEEPAA
	301	PDAWVVEPPE	VPEVAVPEID	ILPPPPVSEI	YNRTYEPPAG	FEQAQRSRIA

351 ETDHLAADVL NGGWQEETAA IADDGSEGAA ERSSGQYLSE TEAFGHDSQA
 401 VCPFEDVPSE RPSCRVSDE ADEGAFQSEE TGAVSEHLPT TDLLLPLPFN
 451 PEATQTEEL LENSITIEEK LAEFKVKVKV VDSYSGPVIT RYEIEPDVGV
 501 RGNVNLNLEK DLARSLGVAS IRVVEIPGK TCMGLELPNP KROMIRLSEI
 551 FNSPEFAESK SKLTALGQD ITGQPVVTDL GKAPHLVAG TTGSGKSVGV
 601 NAMILSMLFK AAPEDVRMIM IDPKMLELSI YEGITHLLAP VVTDMLAAN
 651 ALNWCVNEME KRYRLMSFMG VRNLAGFNQK IAEAAARGEK IGNPFSLTPD
 701 DPEPLEKLPF IVVVVDEFAD LMMTAGKKIE ELIARLAQKA RAAGIHLILA
 751 TQRPSVDVIT GLIKANIPTR IAFQVSSKID SRTILDQMG AENLLGQDML
 801 FLPPGTAYPQ RVHGAFASDE EVHRVVEYLK QFGEPTYVDD ILSGGGSEEL
 851 PGIGRSGDGE TDPMYDEAVS VVLKTRKASI SGVQRALRIG YNRAARLIDQ
 901 MEAEGIVSAP EHNGNRTILV PLDNA*

This partial gonococcal sequence contains a predicted transmembrane region and a predicted
 ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain
 homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK
 (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng: 467 IEKLAEFKVKVKVDSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 526
 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
 FtsK: 868 VEARLADRIKADVVNYSPPGVITRFEINLAPGVKAARISNLSRDLARSLSTVAVRVVEV 927
 ORF58ng: 527 IPGKTCMGLELPNPKRQMIKRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLKGAPHL 586
 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
 FtsK: 928 IPGKPYVGLELPNKKRQTVYLREVLDAKFRDNPSPLTVVLGKDIAGEPVVADLAKMPHL 987
 ORF58ng: 587 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDML 646
 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDML
 FtsK: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLETVVTDML 1047
 ORF58ng: 647 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 704
 AANAL WCVNEME+RY+LMS +GVRNLG+N+KIAEA I +P+ D +
 FtsK: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPPIPDYWKPGDSMDAQH 1107
 ORF58ng: 705 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 762
 L+K P+IVV+VDEFADLMMT GKK+EEELIARLAQKARAAGIHL+LATQRPSVDVITGL
 FtsK: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167
 ORF58ng: 763 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQDMLFLPPGTAYPQRVHGAFASDEEV 822
 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
 FtsK: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227
 ORF58ng: 823 HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 882
 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
 FtsK: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAELDPLFDQAVQFVTEKRKASISG 1286
 ORF58ng: 883 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 921
 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
 FtsK: 1287 VQRQFRIGYNRAARIEQMEAQIVSEQHNGNREVLAP 1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1 ATGTTTTGGA TAGTTTTGAT CGTTATtgtg TTGCTTGCGC TTGCCGGCCT
 51 GTTTTTTGTG CGCGCACAAAT CCGAACGCGA GTGGATGCGC GAGGTTTCTG
 101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
 151 GACGGTATGC CCGATTTTCC CGAGTTTTCG CTGATGCTTT TCCATGCCGT
 201 CAAAACGGCA GTGTATTGGC TGTTTGTCGG TGTCGTCCGT TTCTGCCGAA
 251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
 301 GCAAACCGTG CGGATGTTCG GACCGCATCC GACGGGTATT CAGACAGTGG
 351 AAACGGGACG GAAGAAGCGG AAACGGAAGC AGCAGAAGCT CGGAGGAAG
 401 AGGCTGCCGgA TACgGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
 451 ATCCcatTCG ACCGGAGTAT TGCTGAAGGG TTGATGCAGT CTGAAAGCAA
 501 AACTTCGCCCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
 551 CGCGTGCTTT AAGCAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
 601 GATGCATTTC AGAAAAACGG AACAGCCGTC CCCAAAGTAC GCGTGTCCGA
 651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
 701 AACGCACGTA TTCCCGTATG TTTGATGCGG ACAAAGAAGC GTTTTCCGAG
 751 TCTGCGGATT ACGGATTTGA GCCGATTTT GAGAAGCAGC ATCCGTCTGC

801 CTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCGCCGTC
 851 ATGCGAGGCA GGAGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
 901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCCGCCGCCG
 951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
 1001 CGCGGATTTT GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAAACGG
 1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAACCGTTTC
 1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAAGCTGCC GATATCCATA
 1151 TTGAAGAGCC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
 1201 CCGGAGGTAG CCGTACCCGA AATCGATATT CTGCCGCCGC CTCCCGTATC
 1251 GGAAATCTAC AACCGTACCT ATGAGCCGCC GGCAGGATTC GAGCAGGCGC
 1301 AACGCAGCCG CATTGCCGAA ACCGACCATC TTGCCGCTGA TGTTTTGAAT
 1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCAGATGACG GCAGTGAGGG
 1401 TGCGGCAGAG CCGTCAAGCG GGCAATATCT GTCGGAACCC GAAGCGTTCG
 1451 GGCATGACAG TCAGGCGGTT TGTCCGTTTG AAGATGTGCC GTCTGAACGC
 1501 CCGTCTGCC GGGTATCGGA TACGGAAGCG GATGAAGGGG CGTTCCAATC
 1551 GGAAGAGACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
 1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAAGCTTTG
 1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
 1701 CCGCTCTGTC GATTCTTATT CCGGCCCGCT GATTACGCGT TATGAAATCG
 1751 AACCCGATGT CGGCGTGC GC AATTCCG TTCTGAATTT GGA AAAAGAC
 1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCTG AAACCATCCC
 1851 CGGCAAAACC TGCATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA
 1901 TACGCTGAG CGAAATTTTC AATTGCGCCG AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATT ACCGGACAGC CCGTCGTAAC
 2001 CGACTTGGGC AAAGCACC GC ATTTGCTGGT TGCCGCGACG ACCGGTTCGG
 2051 GCAAAATCGT GGGTGTCAAC GCGATGATT TGTCTATGCT TTTCAAAGCC
 2101 GCGCCGGAAG ACGTGCATG GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTTAC GAAGGCATCA CGCAGCTGCT CGCCCTGTCT GTTACCATA
 2201 TGAAGCTGGC GGCAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGCAATCTTG CGGGCTTCAA
 2301 CCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
 2351 TCAGCTCAC GCCGACGAT CCCGAACCTT TGGAAAAACT GCCGTTTATC
 2401 GTGGTCTGG TCGATGAGTT TGCCGATTG ATGATGACG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCGC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
 2501 TCCACCTTAT CCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCAGCAA
 2601 AATCGACAGC CGCAGGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGGCGA TATGCTGTTC CTGCCGCGG GTACTGCCTA TCCGACGCGC
 2701 GTTACGCGC GCTTTGCCTC GGATGAAGAG GTGCACGCG TGGTCAATA
 2751 TCTGAAGCAG TTTGGCGAGC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAGCGA AGAGCTGCC GGCATCGGGC GCAGCGGCGA CGGCGAAACC
 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
 2901 CAGCATTTTC GGCGTACAGC GCGCCTTGCG CATCGGCTAC AACCGCGCCG
 2951 CGCGTCTGAT TGACCAATG GAAGCGGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTGGACAATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

1 MFWIVLIVIV LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
 51 DGMPDFPEFS LMLFHAVKTA VYWLFVGVVR FCRNYLAHES EPDRPVPPAS
 101 ANRADVPTAS DGYSDSNGT EEAETEAAEA AEEEEADTED IATAVIDNRR
 151 IPFDRSIAEG LMQSESKTSP VRPVFKEITL EEATRALSSA ALRETKKRYI
 201 DAFEKNGTAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE
 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQEKQ QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTETVSSVG YGGPVYDEAA DIHIEEPAAP DAWVVEPPEV
 401 PEVAVPEIDI LPPPPVSEIY NRTYEPPAGF EQAQRSRIAE TDHLAADVLN
 451 GGWQEETA AI ADDGSEGAAE RSSGQYLSET EAFGHDSQAV CFFEDVP SER
 501 PSCRVSDETA DEGAQSEET GAVSEHLPTT DLLLPLLFNP EATQTEELL
 551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 651 KLTLLALGQDI TGQPVVTDLG KAPHLVAGT TSGSGSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGITHLLAPV VTDMLAANA LNWCVNEMEK
 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPLKLPFI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LKANIPTRI AFQVSKIDS RTILDQMGAE NLLGQGDMLF LPPGTAYPQR
 901 VHGAFAFASDEE VHRVVEYLKQ FGEPTYVDDI LSGGGSEELP GIGRSGDGET
 951 DPMYDEAVSV VLKTRKASIS GVQRLRIGY NRAARLIDQM EAEGIVSAPE
 1001 HNGNRTILVP LDNA*

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALLAGLFFVRAQSEREWREVSAAWQEKKGKQAELEIKDGMDFPELA					
5	orf58ng-1	MFWIVLIVIVLLALLAGLFFVRAQSEREWREVSAAWQEKKGKQAELEIKDGMDFPEFS					
		10	20	30	40	50	60
	orf58-1.pep	70	80	90	100	110	120
10	orf58ng-1	LMLFHAVKTAVYWLFVGVVRFRCNYLAHESEPDPRVPPASANRADVPTASDGYSDSGNGT					
		70	80	90	100	110	120
	orf58-1.pep	130	140	150	160	170	180
15	orf58ng-1	EEAETEEAEAEAEAAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPVPRPVFKEITL					
		130	140	150	160	170	180
	orf58-1.pep	190	200	210	220	230	240
20	orf58ng-1	EEATRALNSAALRETKKRYIDAFKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSHM					
		190	200	210	220	230	240
	orf58-1.pep	250	260	270	280	290	300
25	orf58ng-1	FDADKEAFSESADYGFEYPYFEKQHPSAFSAVKAENARNAPFRRHAGQKGQAEAKSPDVS					
30		250	260	270	280	290	300
	orf58-1.pep	310	320	330	340	350	360
35	orf58ng-1	QQQSVSDGTAVRDARRRVSVNLKEPNKATVSAEARISRLIPESQTVVGKRDVEMPSETEN					
		310	320	330	340	350	360
	orf58-1.pep	370	380	390	400	410	420
40	orf58ng-1	VFTETVSSVGYGGPVYDETAADIHIEEPAAPDAWVPEPPEVVKVPMTAIDIQPPPPVSEIY					
		370	380	390	400	410	420
	orf58-1.pep	430	440	450	460	470	480
45	orf58ng-1	NRTYEPPSGFEQVQRSRIAETDHLADDVLNGGWQEETAADIADDGSEGAERSSGQYLSET					
		430	440	450	460	470	480
	orf58-1.pep	490	500	510	520	530	540
50	orf58ng-1	EAFGHDSQAVCPFENVPSERPSCRVSDEADGAFQSEETGAVSEHLPTDLLLLPPLFNP					
		490	500	510	520	530	540
	orf58-1.pep	550	560	570	580	590	600
55	orf58ng-1	EATQTEEELLENSTIEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVRGNSVLNLEKD					
60		550	560	570	580	590	600
	orf58-1.pep	610	620	630	640	650	660
65	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLLPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
	orf58-1.pep	670	680	690	700	710	720
70	orf58ng-1	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLPPGTAYPQR					
15	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLPPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGE PDYVDDILSGGGSEELPGIGRSGDDETDPMYDEAVSV					
20	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGE PDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
25	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

35	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329 Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
40	Query: 556 IEEKLAEFKVKVVDVSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE Sbjct: 868 VEARLADFRIKADVVNYSPPGVITRFEENLAPGVKAARISNLSRDLARSLSTVAVRVVEV 927
45	Query: 616 IPGKTCMGLLEPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL Sbjct: 928 IPGKPYVGLLEPNKKRQTVYLREVLDAKFRDNPSPLTVVLGKDIAAGEPVVADLAKMPHL 987
50	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLTEVVTDK 1047
55	Query: 736 LAANALNWCNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTTPDDPEP-- 793 AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA I +P+ D + Sbjct: 1048 DAANALRWCNEMERRYKLSALGVRNLAGYNEKIAEADRMMPIDPDYWKPGDSMDAQH 1107
60	Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851 L+K P+IVV+VDEFADLMMT GKK+EEELIARLAQKARAAGIHL+LATQRPSVDVITGL Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLATQRPSVDVITGL 1167
65	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLQGQDMLFLPPGTAYPQRVHGAFASDEEV 911 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDM+ P + P RVHGAF D+EV Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227
	Query: 912 HRVVEYLKQFGE PDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG Sbjct: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286
	Query: 972 VQRALRIGYNRAARLIDQMEAEIGVSAPEHNGNRTILVP 1010 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
     101  TGTCTGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
     151  GCATTGGTCG GCTTCTGGGT C.....
//
10     901  .....A TTGCCATCGG TTTGTTTTTA ATTTACCAAA ACGGGCTGAC
      951  CCTGCTTTT GAAGCCGTGG AAGACGGCAA AATCCATTTT TGGCTCGGAC
    1001  TGTGCGCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTTGCGC
    1051  GTCCGCAGTA TGCCAGCCA GCCCTTCTGG CAGGCGGTTG GCAAAAGTCT
    1101  GACATTGAAA GGCGGAAAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
     51  ALVGFVW... ..
//
    301  ...IAIGLFL IYQNLTLLE EAVEDGKIHF WLGLLPMHII MFVLALILLR
    351  VRSMPSQPFW QAVGKSLTLK GKG*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
     51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
    101  TGTCTGGCCG TGCCGCCGAC GGGCGTGTGC CCATCGATGC CGTGTGCGCA
    151  TTGGTCCGGT TCTGGGTCAT CGGTATGACG CCCTTTTTCG TGGTGTGAC
    201  CGCATTTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CCGGACAGCG
    251  AAATGTCCGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
    301  CCGGTGATGC AGTTGCGGT GCCGTTTGCC GTTTTGGTTG CCGTCATGCA
    351  GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
    401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
    451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTCGAAA CCTTCGATAC
    501  CGAATCCGCG ATCATGAAA ACCTGTTCCT GCGCGAACAG GACAAAAACG
    551  GCGGCGACAA CATCATCTTC GCCAAGAAG GTAACCTCTC GCTGAACGAC
    601  AACAAACGCA CGCTCGAATT GCGCCACGCG TACCGTTACA GCGGCACGCC
    651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAAACCT AACCTGATTA
    701  TCAGCACCAC GCCCAAACCT ATCGACCCCG TTTCCACCG CCGTACCATT
    751  CCGACCGCCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
    801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCTGCTTG
    851  CCGTGCCGCT TTCCTATTTC AACCCGCGCA GCGGACATAC CTACAATATC
    901  TTGATTGCCA TCGGTTTGT TTTAATTAC CAAAACGGGC TGACCCTGCT
    951  TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
   1001  CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
   1051  AGTATGCCCA GCCAGCCCTT CTGCGAGGCG GTTGGCAAAA GTCTGACATT
   1101  GAAAGGCGGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
     51  LVGFVWIGMT PLLVLITAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
    101  PVMQFAVPFA VLVAVMQLWV IPWAEIERSRE YAEILKQKQE LSLVEAGEFN
    151  SLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNGGDNII F AKEGNFSLND
    201  NKRTLELRHG YRYSGTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
    251  PTAQLIGSSN PQHQAELMWR ISLTVSVLLL CLLAVPLSYF NPRSHTYNI
    301  LIAIGLFLIY QNLTLLEFA VEDGKIHFVW GLLPMHIIMF AVALILLRVR
    351  SMPSPQFWQA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

		10	20	30	40	50
5	orf101.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX				
	orf101a	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXADXRX-AIDAVLALVGFVXXM				
		10	20	30	40	50
10	orf101.pepIAIGLFLIYQNGLTLLFEAVEDGKIHFVWGL				
	orf101a	LTVSVLLCLLAVPLSYFNPRSGHTYNILXAIGLFLIYQNGLTLLFEAVEDGKIHFVWGL				
		280	290	300	310	320
15	orf101.pep	120	130	140	150	
	orf101a	LPMHIIMFVLALILLRVRSMPSQPFQAVGKSLTLKGGKX				
20	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFQAVGKSLTLKGGKX				
		340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

	1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
	51	CATTTTCGTC	GTCCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
	101	TGCTCGGCCN	TGCCGCCGAC	NGGCGTNTCG	CCATCGATGC	CGTGTGGCA
25	151	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
	201	CGCATTTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CGNGACAGCG
	251	AAATGTCGGT	CTGGNTATCC	TGCGGATTGG	CATTGAAACA	ATGGATACGC
	301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGTGGTTG	CCGTCATGCA
30	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
	401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAGGCAGG	CGGGTTC AAC
	451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTTGTGCGAA	CCTTCGATAC
	501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCTC	GCGCGAACAG	GACAAAAACG
	551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTTCT	GCTGAACGAC
35	601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
	651	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAACCT	AACCTGATTA
	701	TCAGCACCAC	GCCCAAACTC	ATCGACCCCG	TTTCCCACCG	CCGTACNATN
	751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
	801	GATGTGGCGC	ATCTCGCTGA	CCGTCAGCGT	CCTCCTACTC	TGCCTGCTTG
	851	CCGTGCGCGT	TTCTTATTTT	AACCCGCGCA	GCGGACATAC	CTACAATATC
40	901	TTGANTGCCA	TCGGTTTGTT	TTTAATTTAC	CAAAACGGGC	TGACCCTGCT
	951	TTTTGAAGCC	GTGGAAGACG	GCAAAATCCA	TTTTTGGGTC	GGACTGCTGC
	1001	CTATGCACAT	CATCATGTTT	GTCATCGCAA	TCGTACTTCT	GCGCGTCCGC
	1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
	1101	GAAAGGCGGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
	51	LVGFVWXXMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
	101	PVMQFAVPFA	VLVAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGGFN
50	151	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSLND
	201	NKRTLELRHG	YRYSGTPGRA	DYNQVSFXKL	NLIISTTPKL	IDPVSHRRTX
	251	PTAQLIGSSN	PQHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LXAIGLFLIY	QNGLTLLFEA	VEDGKIHFVW	GLLPMHIIME	VIAIVLLRVR
	351	SMPSQPFQQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXADXRXAIDAVLALVGFVWXXMT	60
	orf101-1	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT	60
	orf101a.pep	PLLLVLTAFISTLTVLTRYWRDSEMSVWXSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120
60	orf101-1	PLLLVLTAFISTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120

	orf101a.pep	IPWAEILRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
5	orf101-1	IPWAEILRSREYAEILKQKQELSLVEAGEFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
	orf101a.pep	DKNGGDNIIFXKESNFSLNDNKRTLELRHGYRYSPTGRADYNQVSFXKLNLIISTTPKL	240
	orf101-1	DKNGGDNIIFAKEGNFSLNDNKRTLELRHGYRYSPTGRADYNQVSFQKLNLIISTTPKL	240
10	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHXAELMWRLSLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHXAELMWRLSLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
15	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFGLLPMHIIMFAVALILLRVRSMPSQPFWQA	360
	orf101a.pep	VGKSLTLKGGK	371
20	orf101-1	VGKSLTLKGGK	371

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

gonorrhoeae:

	orf101.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVW	57
	orf101ng	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFVWIGM	59
30		//	
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHFGLG	333
35	orf101ng	SLTVSVLLCLLAVPLSYFNPRSGHTYNI	331
	orf101.pep	LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK	373
	orf101ng	LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial

amino acid sequence <SEQ ID 504>:

	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGEFVIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVPFA	ILIAVMQLWV	IPWAEILRSRE	YAEILKQKQE	LSLVEAGEFN
45	151	NLGKRNGRVY	FVETFDTEG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLKD
	201	NKRTLELRHG	YRYSPTGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFGL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
	51	CATTTTCGTC	GTCTCTTGG	CGGTGTTGGT	GTCCACGCAG	GCGATCAACC
	101	TGCTTGGCCG	CGCAGCTGAC	GGGCGTGTCG	CCATCGATGC	CGTGTGGCC
	151	TTAGTCGGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
	201	CGCATTCATC	AGCACGCTGA	CCGTATTGAC	CCGCTACTGG	CGCGACAGCG
55	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTTGAAACA	GTGGATACGC
	301	CCCGTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATTG	CCGTCATGCA
	351	GCTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAAGCCGG	CGAGTTCAAT
	451	AACTTGGGCA	AGCGCAACGG	CagggtttaT	TtcgtcgaaA	CCTTTGACAC
60	501	CGaatccgGC	ATCATGAAAA	ACCTGTtcct	GcGCGAACAG	GACAAAAACG
	551	gcggcgcacaA	CATCATCTTC	GCcaaaGAag	gtaactTctc	gctgaaggac

10

20

25

30

35

40

45

50

55

60

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTCTGC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
    10  251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
     301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPOYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWQDVRV AGQNDVAATG DAHSPILNNA AANTSNTAN
     101  NGTHIPLFAI DTGKLGGXVC QQNHLDQYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTAKPOYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      GGG INA+ TLT+ P   G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspa    GGGLINASVTLTSGVPLNNGNLTGFDVSSGKVVIIGGKGLDTSADADYTRILSRAEINA 256

25      orf113  PVWQDVRVAGQNDVAATGDAHSPIILXXXXXXXXXXXXXGTHIPLFAIDTGKLGGMYA 120
      VWG+DV+VV+G+N + G               + P AIDT LGGMYA
      pspa    GVWGRDVKVVSQKKNLDFDG-----SLAKTASAPSSSDSVTPTVAIDTATLGMYA 307

30      orf113  NKITLISTVEQAGIRNQGFASAGNAVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      +KITLIST  A IRN+G+ FA+ G V ++A+GKL N+G I A      +++ A+ V N
      pspa    DKITLISTDNGAVIRNKGRIFAATGGVTLSDAGKLSNSGSIDAA----EITISAQTVDN 362

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113                                GGGFINASCATLTAKPOYQAGDLSAFKIR 30
      ||||| |||||:|||||:|||||
      orf113ng  SHPSQLNGYIEVGRRAEVVIANPAGIAVNGGGFINASRATLTGQFPYQAGDFSGFKIR 224

40      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWQDVRVAGQNDVAATGDAHSPILNNA 90
      |||:|||||:|||||:|||||
      orf113ng  QGNAVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263

      orf113                                IDTGKLGGXVCQQNHLDQYGRASRHS 135
      ||||| |||||:|||||
45      orf113ng  DFGFKIRQGNVVIAGHGLDARDTDFTRILVCQQNHLDQYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLRYVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
      51  SKAFCF SALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNIP

```

-300-

101 QVNIQTPTSA GVSVNQYQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
 151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
 201 ASRATLTGQ PQYQAGDFSG FKIRQNAVI AGHGLDARDT DFTRILVCQQ
 251 NHLDQYGRS RHS*

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1  ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51  CAACATTTCAT CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
      101  GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
      151  TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
      201  ATACATTATC AATCCTGTCA ATAAAGGCTA TCTGTTGAA ACCGATCCAC
      251  GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGACAGC
      15  301  CTCAAACCTAG ACCCAAACAA TTACATAAA CGTTTGGGTG ATGGTTATTA
      351  CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
      401  GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
      451  AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
      501  AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
      20  551  AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
      601  CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
      651  GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
      701  CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
      751  GACAATATCG GTGGGCGGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
      25  801  ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
      851  TGCTCAACGC AGGCAACAAAC ATCAACAGCC AAAGCACCAC CGCCAGCACT
      901  CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
      951  TATCACAGGC AAAGAAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1  ..STGHSEQNYT LPREITRNI LGSFAYESHR KALSHHAPSQ GTELPQSNGI
      51  SLPYTSNSFT PLPSSSLYII NPVNKGylVE TDPRFANYRQ WLGS DYMLDS
      101  LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
      151  NGATAARSMN LSVGIALSAE QVAQLTSDIV WLQKEVKLP DGGTQTVLVP
      201  QVYVRVKN GD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
      35  251  DNIGGRIHAQ KSAVTATQDI NNIGMMLSAE QTLNAGNN INSQSTTASS
      301  QNTQGSSTYL DRMAGIYITG KEKGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

```

40      Orf115: 1    STGHSEQNYTLPREITRNI LGSFAYESHRKALSHHAPSQGTLPQSNGISLPYTSNSFT 60
      STG+S  Y    E++ +I +G AY+ +  +  P    +  NGI  +T
      pspA:  778  STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQASDIPGTVPVVAENGIHPTFT----- 831

      Orf115: 61    PLPSSSLYIINPVNKGylVETDPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQR 120
      LP+SSL+ I P  NKGyl+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+
      45  pspA:  832  -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSGYMLAALQDPNHIHKRLGDGYEQK 890

      Orf115: 121   LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
      L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
      50  pspA:  891  LVNEQIAKLTYRRLDGYTNDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIV 950

      Orf115: 181   WLQKEVKLPDGGTQTVLVPQVYVRVKN GDIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
      WL  + V LPDG TQTVL P+VYVR +  D++G+GALLSGS  I  SG+++N  G IAG
      55  pspA:  951  WLENETVTLPDGTTQTVLKPVKYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009

```

-301-

Orf115: 240 RNALIINTDTLDNIGGRIHAQSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXXXX 299
 R ALI+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNIKNLQGDQGNIFAAAGSDITNTGS-IGAENALLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXYLDRMAGIYITGKEKG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorrhoeae*:

	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
		: :	
15	orf115ng	NEQTFGEKKVFSENGKLHNYWRARRKGHDETHREQNYTLPEEITRDISLGSFAYESHK	71
	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGYLVE	81
		: :	
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET	131
20	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYYEQRLINEQIAELTGHRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLQKEVKLPDGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNQDIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
30			
	orf115ng	VYVRVKNQDIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
35	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
	orf115.pep	EKG	325
	orf115ng	EKGVLAAQAGKDINIIAGQISNQSDQGTQLQAGRDINLDTVQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino
 40 acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGSDYMLGS
45	151	LKLDPNNLHK	RLGDGYYEQ	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
	251	QYVRVKNKG	IDGKGALLSG	SNTQINVS	GLKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGT
50	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTKGDVTLL
	451	SGNNLNAAK	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GKNLVITDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTOKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMDIGAAQ
55	651	NQLNSKTTQT	YEOKGLTVAF	SSPVTDLAQ	AIATAHKA	QFDKAKTTAL
	701	MPWRLPMQVG	RLFQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTGACACAAC	TACTGGCGTG
60	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTATACT
	151	TGCGCCGAGG	AAATCACACG	CGACATTTC	CTGGGTTTCA	TTGCCTATGA
	201	ATCGCATAGC	AAAGCATTA	GCCGTCATGC	GCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTGCGAAAAG	CAACGGTATT

-302-

301 TCGCTACCTT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
 451 CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
 601 AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
 701 AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
 751 CAGGTTTATG TACGCGTTAA AATGGCGGCG ATAGACGGTA AAGGTGCATT
 801 GTTGTGAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
 851 CAGGCACGAT TGCAAGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
 901 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
 1001 TGGTCAATGC GGGTAACAAC ATCAACAACC AAAGCAGCGC CAAGAGCAGT
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAGAGACA
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTCG GATACGGTAC AAACCGGCAA
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCtatTG
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAGG
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
 1651 CAAGCAGGCA ATCATGTTTC CATTTGGTACA ACCCAAACCTC AAAGCCAAAG
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
 1851 TGTGTCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
 2001 GGTGGCATTG AGTTCGCCCC TTACCGATTT GGCACAACAA GCGATTGCCG
 2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
 2151 GCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
 51 LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPHFANYRQ WLGS DYMLGS
 151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
 45 251 QVYVRVKNNG IDKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLLNAGNN INNQSTAKSS
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINI IAGQ ISNQSDQGT
 401 RLQAGRDINL DTVQTKGYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
 451 SGNLNAKAA EVGSAKGLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 50 501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGT
 551 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ
 651 NQLNSKTTQT YEQKGLTFAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
 701 MPWRLPMQVG RPIKQAKAHK T*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

	20	30	40	50	60	70
orf115ng-1.p	NEQTFGEKKV	VFSENGKLHNY	YWRARRKGHD	ETGHREQNYTL	LPEEITRDIS	LGSFAYESHK
				: :		
orf115				STGHSEQNYTL	PREITRNIS	LGSFAYESHK
				10	20	30
	80	90	100	110	120	130
orf115ng-1.p	ALSRHAPSQ	GTLPQSNRD	NIRTAKSNGI	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVET
	:					
orf115	ALSHHAPSQ	GTLPQSN	-----	GISLPYTSN	SFTPLPSS	SLYIINPVN
	40	50	60	70	80	

-303-

		140	150	160	170	180	190
5	orf115ng-1.p	DPRFANYRQWLGS	SDYMLGSLKLD	PNNLHKRLGD	GYEQRLINEQIA	ELTGHRRLDGY	QND
	orf115	DPRFANYRQWLGS	SDYMLDSLKLD	PNNLHKRLGD	GYEQRLINEQIA	ELTGHRRLDGY	QND
		90	100	110	120	130	140
10	orf115ng-1.p	EEQFKALMDNGATA	AARSMNSVGIAL	SAEQAAQLTSDI	VWLQKEVKLPD	GGTQTVLMPQ	
	orf115	EEQFKALMDNGATA	AARSMNSVGIAL	SAEQAAQLTSDI	VWLQKEVKLPD	GGTQTVLMPQ	
		150	160	170	180	190	200
15	orf115ng-1.p	VYVRVKNGGIDGK	GALLSGSNTQIN	VSGSLKNSGTI	AGRNALIINTD	TLDNIGGRIHAQ	K
	orf115	VYVRVKNGGIDGK	GALLSGSNTQIN	VSGSLKNSGTI	AGRNALIINTD	TLDNIGGRIHAQ	K
		210	220	230	240	250	260
20	orf115ng-1.p	SAVTATQDINNIG	GILSAEQTLN	NAGNNINNQS	TAKSSQNAQGS	STYLDRMAGIY	ITGK
	orf115	SAVTATQDINNIG	GMLSAEQTLN	NAGNNINSQ	STTASSQNTQ	GSSTYLDRMAGI	YITGK
		270	280	290	300	310	320
25	orf115ng-1.p	EKGVLAQAQAKD	INIIAGQISNQ	SDQGQTRLQAG	RDLNLDTVQT	GKYQEIHFDA	NHTIR
	orf115	EKGV					

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30	gi 2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273 Score = 604 bits (1541), Expect = e-172 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
35	Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I Sbjct: 739 LIVGTPESALDNDDELTKTKTI-TDKGDLHRYHRHHKKGRDSTGYRSRSPYEPAPEVS-SIR 796
40	Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840
45	Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLGD GYEQRLINEQIAELT 180 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGD GYEQ+L+NEQIA+LT Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHIHKRLGD GYEQKL VNEQIAKLT 900
50	Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNSVGIALSAEQAAQLTSDIVWLQKEVKLP 240 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTLP 960
55	Query: 241 DGGTQTVLMPQVYVRVKNGGIDGKALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N Sbjct: 961 DGT TQTVLKP KYVVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019
60	Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQS TAKSSQNAQGSSTY 359 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS Sbjct: 1020 IKNLQGLDQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078
65	Query: 360 LDRMAGIYITGKEKGVLAQAQAKDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138
	Query: 420 EIHFDADNHTIRGSTNEVGSSIQTGKDVTLLSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479 FD+DN+ IR NEVGSI+T+G+++L + ++ +AAEVGS +G L + A DI + Sbjct: 1139 NTIFSDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198
	Query: 480 SSGIHAGQVDDASKHTGRSGGKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

Query: 540 SNVISDNGTRIQAAGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAATRSRRAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYYTQTGSTISSPQGDVGISGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT D 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTVVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

```

1  ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
51  TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
20 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
251 ATGCCAACAT CCTTGGCAGC AATGTTATT CCGATAATGG CACCCAGATT
301 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
401 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAAGC
25 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
501 TGTTGCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
651 GGTGGCATTC AGTTCGCCCG TTACCGATT GGCACAACAA ...

```

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

```

1  ..SGNNLNAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
51  GNKLVIDTKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTQI
101 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
151 NEHTGSTVGS LKGDTTIVAG KHYEQIGSTV SSPEGNNIY AQSIDIAAH
35 201 NKLNSNTTQT YEQKXLTVAE SSPVTDLAQQ ...

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

40 Orf117: 4 NLNAKAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLVIDTKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
 pspA: 1173 DIRIRAAEVGSEQRLKLAAGRDIVKEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232

45 Orf117: 64 HETAQSSTFEKGQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAATRSRRAEMNKK 1292

50 Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGSGLKGDTTIVAGKHYEQIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRETSTVSHTESVVGSLNGNTLISAGKHYYTQTGSTISS 1352

Orf117: 183 PEGNNIYQAQSIDIAAHNKLNSNTTQTYEQKXLTVAFSSPVT D 226
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
 pspA: 1353 PQGDVGISGKISIDAAQNRYSQESKQVYEQKGVTVVAISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFADNHTIRGSTNEVGSSIQTKGDVTLSSGNNLNAAAEVGSAGKTLAVYAKNDITIS	480
	orf117.pep	AGINTTHVDDASKHTGRSGGKNLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
10	orf117ng	SGIHAGQVDDASKHTGRSGGKNLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSGLMSAGIGFTIGSKTNTQENQSQS	600
15	orf117.pep	NEHTGSTVGSGLKGDTTIVAGKHYEQIGSTVSSPEGNNNTIYAQSIDIQAHNKLNSTTQT	210
	orf117ng	NEHTGSTVGSGLKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMDIGAAQNQLNSKTTQT	660
20	orf117.pep	YEQKXLTVAFFSPVTDLAQQ	230
	orf117ng	YEQKGLTVAFSSPVTDLAQQAI(AVAHKA)AKQFDKAKTTALMPWRLPMQVGR(LFKQAK)APK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LPEEITRDIS LGSFAYESHK KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
	151	LKLDPNNLHK RLG DGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP
30	251	QYVVRVKNNG IDGKGALLSG SNTQINVSQS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
	351	QNAQSSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGT
	401	RLQAGRDINL DTVQTGYQE IHFADNHTI RGSTNEVGSS IQTKGDVTL
	451	SGNNLNAAAEVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
35	501	GNKLVIDKA QSHHETAQSS TFEKQVVLQ AGNDANILGS NVISDNGT
	551	QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKA AKQFDKAKTTAL
	701	MPWRLPMQVG RLFKQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTGTGTC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGACAAAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATATACT
	151	TTGCCGGAGG AAATCACACG CGACATTTCAT CTGGGTTTCAT TTGCCTATGA
45	201	ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCCAGCCAA GGCCTGAGT
	251	TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
	301	TCGCTACCCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGAA ACCGATCCAC
	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
50	451	CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGTTATTA
	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTAAAGC CTTAATGGAT
	601	ATGGCGCGCA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
55	701	AAAAAGAAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
	801	GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT
	851	CAGGACGAT TGCAGGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
	901	GACCAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
60	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGS CAGGTATTTA

1101 TATCACAGGC AAAGAAAAAG GTGTTTATAG AGCGCAGGCA GGCAAAGACA
 1151 TCAACATCAT TGCCGTCAG ATCAGCAATC AATCAGATCA AGGGCAAACC
 1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAG
 5 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCCTattTG
 1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
 10 1551 TCAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 1601 AATGCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
 1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
 1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC
 15 1801 AACGCAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
 1851 TGTTCGAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
 2001 GGTGCACTTC AGTTGCCCGG TTACCGATTG GGCACAACAA GCGATTGCCG
 20 2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
 2151 GGCACACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

1 LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
 25 51 LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGS DYMLGS
 151 LKLDPNNLHK RLGDGYEYQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
 251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 30 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
 351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQGQT
 401 RLQAGRDINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
 451 SGNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 501 GNKLVIDTKA QSHHETAQSS TFEGQVVLQ AGNDANILGS NVISDNTRI
 35 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS
 601 NEHTGSTVGS LKGDPTIVAS KHYEQTGSNV SSPEGNNLIS TQSMDIGAAQ
 651 NQLNSKTTQT YEQKGLTVAE SSPVTDLAQ AIAVAHKAAN KSDKAKTTAL
 701 MPWRLEPMQVG RPIKQAKAHK T*

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it

40 shows homology with a secreted *N.meningitidis* protein in the database:

gi|2623258 (AF030941) putative secreted protein [*Neisseria meningitidis*]Length = 2273

Score = 604 bits (1541), Expect = e-172

Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

45 Query: 1 LLVQTEKDGLHNEQTFGEKKVFSENGKLHNYWRARRKGHDETGHREQNYTLPEEITRDIS 60
 L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
 Sbjct: 739 LIVGTPESALDNDDELGTCTI-TDKGDLHRYHRHHKGRDSTGYRSRSPYEPAPEVS-SIR 796
 50 Query: 61 LGSFAYESHKALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
 Sbjct: 797 MGISAYKGY-----APQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840
 55 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEYQRLINEQIAELT 180
 P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEYQ+L+NEQIA+LT
 Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGSGYMLAALQQDPNHIHKRLGDGYEYQKLVEQIAKLT 900
 60 Query: 181 GHRRLDGYQNDDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLQKEVKLP 240
 G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
 Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQLTGIALSAEQVARLTSDIVWLENETVTL 960
 Query: 241 DGGTQTVLMPQVYVRVKNNGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
 DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
 Sbjct: 961 DGTQTVLKPVKYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019
 65 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNINNQSTAKSSQNAQGSSTY 359

-307-

+ N+ G + + A D I N G I A E L L L A N N I ++S +S+QN QGS
 Sbjct: 1020 IKNLQGDLOQGNIFAAAGSDITNTGSI-GAENALLKASNNIESRSETRSNQNEQGSVRN 1078
 Query: 360 LDRMAGIYITGKEKGVLAAGKDNIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQs+ GQT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSDTTGISRNQ 1138
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVITLLSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479
 FD+DN+ IR NEVGs+I+T+G+++L + ++ +AAEVGS +G L + A DI +
 Sbjct: 1139 NTIFDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKV 1198
 Query: 480 SSGIHAGQVDDASKHTGRSGGKNLVIDTKAQSHHETAQSSTFEGKQVVLQAGNDANILG 539
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEILVSGRDITVTG 1258
 Query: 540 SNVISDNTRIQAQGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNIVLKAATRSRRAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRs 1318
 Query: 599 QSNEHTGSTVGSLSKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISGKISIDAAQNRYsQESK 1378
 Query: 659 QTYEQKGLTVAFFSPVTD 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTV AISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAwAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACC GGC
 35 201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGyCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA
 40 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CGTGGCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKVR DQFGHSDKDA LLNSXTSHVR
 51 DGKPSGGSV MPKPQFAVKK TAKPDQPMR NLQEODAVYI AKQKQAKASP
 45 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCG
 50 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCCAACC GGC
 201 GGTCAAAAAA ACGGCAAAAC CCCAAGACCC CGCCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGGA GAAAGCGGCA TTATCGGCAA
 55 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCACCGAAAC
 401 CTGCCGACGC GCGGCAAAA CCTGCACCCG TTCCGCAAAC ACCTGCAAAA
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA
 501 CGTGGCCTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGCGC
 60 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGGCATCCG

-309-

5
10
15
20
25
30
35
40
45
50
55
60
65

```

651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTTCGCA
751 CACAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
851 CCATCCATTG GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGGGAA GACGACGGCG CGTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCCTATAA AGGCTTCAGT
1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCAGAAAAA CCTTCGACGA
1101 TTTGTTTATG GATTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
1151 TGGTCAACGA CAAATGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGGC AAAACCGCAT TGGCCTGTT CTCCTAA

```

This encodes a protein having amino acid sequence <SEQ ID 528>:

15
20
25
30
35
40
45
50
55
60
65

```

1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
151 PLITLKELSK VELPWFDFVF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHDFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNPFTNAL LDNQSYKGFS
351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIEPGG KTLRLRFS*

```

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25
30
35
40
45
50
55
60
65

```

                10      20      30      40      50      60
orf119a.pep    MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM
                |||
orf119-1       MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVRDGKPSGGPVM
                |||
                10      20      30      40      50      60
orf119a.pep    MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
                |||
orf119-1       MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH
                |||
                70      80      90      100     110     120
orf119a.pep    TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
orf119-1       TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
                70      80      90      100     110     120
orf119a.pep    TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
orf119-1       TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
                130     140     150     160     170     180
orf119a.pep    TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
orf119-1       TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
                130     140     150     160     170     180
orf119a.pep    TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
orf119-1       TVPEPQTGHSAPKPADAPAKPVPVPQTPAKPLITLKELSKVELPWFDFVDFDISYIALTE
                |||
                190     200     210     220     230     240
orf119a.pep    AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
                |||
orf119-1       AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
                |||
                190     200     210     220     230     240
orf119a.pep    AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
                |||
orf119-1       AKELHALPRLSNRCRYQIVGCTMDHDFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS
                |||
                250     260     270     280     290     300
orf119a.pep    AFNRQVDAFAHSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS
                |||
orf119-1       AFNRQVDAFAHSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS
                |||
                250     260     270     280     290     300
orf119a.pep    AFNRQVDAFAHSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS
                |||
orf119-1       AFNRQVDAFAHSMGGQTLHTDLAAFIIEVASALDAFCARVDQTIAIHLVSPTSISGVELRS
                |||
                310     320     330     340     350     360
orf119a.pep    AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
                |||
orf119-1       AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
                |||
                310     320     330     340     350     360
orf119a.pep    AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
                |||
orf119-1       AVTGVGFVLEDDGAFHYTDTSGSTMFSICSLNNEPFTNALLDNQSYKGFSMLLDIPHSPA
                |||
                370     380     390     400     410     420
orf119a.pep    GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG
                |||
orf119-1       GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG
                |||
                370     380     390     400     410     420
orf119a.pep    GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG
                |||
orf119-1       GEKTFDDLFDLAVRLSGQLNLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG
                |||

```

-308-

5 651 CTATCAGGCA TTTATCGTGG GTATTCAAGC AGTCAGCCGC AACGGACTTG
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTTCGA
 751 CAAAGCATGG GCGGTGAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
 801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACCATCG
 851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
 901 GCCGTAAACG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCACTA
 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
 1001 AGCCGTTTAC CAACGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGGAAAAAA CCTTCGACGA
 1101 TTTGTTTATG GATTGTCGGG TACGCTGTG CCGCCAGTTG AACCTGAATC
 1151 TGGTCAACGA CAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG
 1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
 1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
 51 DGKPSGGSVMPKPKQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
 151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
 20 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
 251 QSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIATHLVSP TSISGVELRS
 301 AVTGVGFVLE DDGAHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFs
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
 401 RTYVVLARQSE MLKVGIEPPG KTALRLFS*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSXTSHVRDGKPSGGSV					
	orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSHVRDGKPSGGPVM					
		10	20	30	40	50	60
35	orf119.pep	MPKQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
	orf119a	MPKQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
		70	80	90	100	110	120
40	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFDVRIDFISY					
	orf119a	TVPEPQTGHSAPKPADAPAKPVPQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE					
		130	140	150	160	170	180
45	orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
		190	200	210	220	230	240

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50 1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
 51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTTCG
 101 GGCACCTCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCGGC
 201 GGTCAAAAAA ACGGCAAAAT CCCAAGACCC CGCCATGCGC AACCTGCAAG
 251 AGCAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 55 301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCCAC ACCGTTCCCG AACCCCAAAC CGGACATTCC GCACCAAAAC
 401 CTGCCGACGC GCCGGCAAAA CCTGTTCCCG TTCCGCAAAAC GCCGGCAAAA
 451 CCGCTGATTA CGCTCAAAGA GCTGTGGAAG GTCGAGCTGC CCTGGTTTGA
 501 CGTGGCGTTC GACTTCATCT CTTATATCGC GCTGACCGAA GCCAAAGAAC
 60 551 TGCACGCACT GCCGCGCCTT TCCAACCGCT GCCGCTACCA GATTGTGCGC
 601 TGCACCATGG ACGACCATT TCCAGATTGCC GAACCCATCC CGGGCATCCG

-310-

```

orf119a.pep    KTALRLFSX
               |||||
orf119-1       KTALRLFSX

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N.gonorrhoeae*:

```

10 orf119.pep    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSXTSHVRDQKPSGGGSM 60
    orf119ng     MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRDQKPSGGPVM 60

15 orf119.pep    MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH 120
    orf119ng     MPKPQPAVKKPAKPQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH 120

    orf119.pep    TVSEPQTGHSATKPADASAKPAPVQTPAKPLITLKELSKVELSWFDVRIDFISY 175
    orf119ng     TVSEPQTGHSAPKPADAPAKPVPVQTPAKPLITLKELSKVELPWFDVRFDFISYIALTE 180

```

The complete length ORF119ng nucleotide sequence <SEQ ID 529> is:

```

20 1 ATGATTTACA TCGTACTGTT CCTCGCCGCC GTCCTCGCCG TTGTCGCCTA
    51 CAATATGTAT CAGGAATAAC AATACCGCAA AAAAGTGCGC GACCAGTTCG
    101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
    151 GACGGCAAAC CGTCCGGCGG GCCAGTCATG ATGCCGAAAC CCCAACCAGC
    201 GGTCAAAAAA CCGGCCAAAC CCCAAGACTC CGCCATGCGC AACCTGCAAG
    25 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
    301 TTCAAACCGG AAATCGAAAC CGCCTTGGA GAAATCGGCA TTATCGGCAA
    351 CTCCGCCAC ACCGTTCCG AACCCCAAAC CGGACATTCC GCACCAGAAC
    401 CTGCGGACGC GCCGGCAAAA CCCGTTCCCG TTCCGCAAA GCGCGCAAAA
    451 CCGCTGATTA CGCTCAAGA GCTGTCGAAG GTCGAGCTGC CCTGGTTTGA
    501 CGTGGCTTc gACTTCATCT CCTATATCGC GCTGACCGAA GCCAAGAAC
    551 TGCACGCACT GCCGCGCCTT tccAACCGCT GCCGCTACCA GATTGTCGGC
    601 TGCACCATGG ACGACCATT CCAGATTGCC GAACCCATCC CGGGCATCCG
    651 CTATCAGGCA TTTATCGTGG GTATCCAGGC AGTCAGCCGC AACGGACTTG
    701 CCTCGCAGGA AGAACTCTCC GCATCAACC GCCAGGCGGA CGCATTCGCA
    35 751 CAAAGCATGG GCGGTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
    801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGGTCGAC CAGACCATCG
    851 CCATCCATTT GGTTCGCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
    901 GCCGTAACGG GCGTGGGTTT CGTTTGGAA GACGACGGCG CGTTCCTACTA
    951 TACCGACACG TCGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
    40 1001 AGCGCTTTAC CAATGCCCTT TTGGACAACC AGTCCTACAA AGGCTTCAGT
    1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GCGCAAAAAA CCTTCGACGA
    1101 TTTGTTTATG GATTGGCGG TACGCTGTG CCGTCAGTTG AACCTGAATC
    1151 TGGTCAACGA CAAATGGAA GAAGTTTGA CCCAATGGCT CAAAGACGTA
    1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAG TCGGTATCGA
    45 1251 ACCGGGCGGC AAAACGCCC TCGCCTGTT TTCATAA

```

This encodes a protein having amino acid sequence <SEQ ID 530>:

```

50 1 MIYIVFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
    51 DGKPSGGPVM MPKPQPAVKK PAKPQDSAMR NLQEQDAVYI AKQKQAKASP
    101 FKTEIETALE EIGIIGNSAH TVSEPQTGHS APKPADAPAK PVPVPQTPAK
    151 PLITLKELSK VELPWFDVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
    201 QTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQADAFa
    251 QSMGQTLHT DLAAFIEVAS ALDAFCARVD QTIAIHLVSP TSISGVELRS
    301 AVTGVGFVLE DDGAFHYTDT SGSTMFISICS LNNEFTNAL LDNQSYKGFS
    351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
    55 401 RTYVLARQSE MLKVGIEPGG KTALRLFS*

```

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

```

60 orf119ng      10 20 30 40 50 60
    MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRDQKPSGGPVM
    |||||:|||||
orf119-1      10 20 30 40 50 60
    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLNSKTSHVDRDQKPSGGGSM

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKPQDSAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEEIGIIGNSAH			
5	orf119-1	MPKPQPAVKKTAKPQDPAMRN	LQEQDAVYIAKQKQAKASPF	KTEIETALEESGIIGNSAH			
		70	80	90	100	110	120
	orf119ng	TVSEPQTGHSAPKPADAPAKP	VPVPQTPAKPLITL	KELSKVELPWF	DVRFDFISYIALTE		
10	orf119-1	TVSEPQTGHSAPKPADAPAKP	VPVPQTPAKPLITL	KELSKVELPWF	DVRFDFISYIALTE		
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPGIRYQAF	IVGIQAVSRNGLASQEELS			
15	orf119-1	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPIPGIRYQAF	IVGIQAVSRNGLASQEELS			
		190	200	210	220	230	240
	orf119ng	AFNRQADAFASMGQT	LHTDLAAFIEVASALDAFC	ARVDQTTIAIHLVSP	TSSISGVELRS		
20	orf119-1	AFNRQVDAFASMGQT	LHTDLAAFIEVASALDAFC	ARVDQTTIAIHLVSP	TSSISGVELRS		
		250	260	270	280	290	300
	orf119ng	AVTGVGVFLEDDGAFHYT	DTSGSTMFSICSLNNEP	FTNALLDNQSYKGF	SMLLDIPHSPA		
25	orf119-1	AVTGVGVFLEDDGAFHYT	DTSGSTMFSICSLNNEP	FTNALLDNQSYKGF	SMLLDIPHSPA		
		310	320	330	340	350	360
	orf119ng	GEKTFDDLFLMDLAVRL	SGQLNLNLVNDKMEEV	STQWLKDVRTYVLAR	QSEMLKVGIEPGG		
30	orf119-1	GEKTFDDLFLMDLAVRL	SGQLNLNLVNDKMEEV	STQWLKDVRTYVLAR	QSEMLKVGIEPGG		
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTTCCTCCA
50	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
	151	CTGGTGTCCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGCGCGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCGGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEDFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVLSAAV
60	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

35 Computer analysis of this amino acid sequence gave the following results:

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

55 orf134.pep ARHGTEDEFMNNSDXIRQIVESTTGTMKLL
|||
orf134a GESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTEDEFMNNSDSIRQIVESTTGTMKLL
210 220 230 240 250 260
40 50 60 70 80 90

-313-

orf134.pep ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
 orf134a ISSIALISLVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
 270 280 290 300 310 320
 5
 orf134.pep LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
 orf134a LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
 10 330 340 350 360 370 380
 orf134.pep LAQDX
 orf134a LAQDX
 15

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

1 ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
 51 GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCATTGG
 101 GCAACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
 151 AACACCATCA GCATCTTCCC AGGGCGCGGC TTCGGCGACA GGCGCAGCGG
 201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
 251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
 301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
 351 TTTGACGCTG CGCGGGCTGA AGCTGGAAC GGGCGGCTG TTTGACGAAA
 401 ACGATGTGAA AGAAGACGCG CAGGTCGTGC TCATCGACCA AAATGTCAAA
 451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAACCA TTTGTTCAG
 501 GAAACGCCCT TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAACGCTT
 551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
 601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
 651 AGACAATGCC AATACCCAGG TTGCGGAAAA AGGGCTGACC GATCTGCTCA
 701 AAGCGCGGCA CGGCACGGAA GATTTCTTCA TGAACAACAG CGACAGCATC
 751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCTCT
 801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
 851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
 901 ATCGGCGCGC GCGCGGCAA TATTTTGAG CAGTTTTGA TTGAGGCGGT
 951 GTTAATCTGC GTCATCGGCG GTTTGGTTCG CGTGGGTTTG TCCGCCGCCG
 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTCC
 1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
 1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGATG
 1151 CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 536>:

1 MSVQAVLAHK MRSLLTMLGI IIGIASVVS VALGNGSQKK ILEDISSIGT
 51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
 151 DKL FADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
 251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
 351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

orf134a.pep MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
 orf134-1 MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG
 55 orf134a.pep FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
 orf134-1 FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
 60 orf134a.pep RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
 orf134-1 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
 orf134a.pep ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
 orf134-1 ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
 65

```

5  orf134a.pep  DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1    DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134a.pep  IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
   orf134-1    IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
10  orf134a.pep  STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1    STGIGIAFGFMPANKAAKLNPIDALAQDX

```

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

```

   orf134.pep  ARHGTEFFMNSDXIRQIVESTTGTMKLL  30
   orf134ng    GESHTNSITVKIKDNANTRVAEKGLELLKARHGTEFFMNSDSIRQIVESTTGTMKLL  264
20  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG  90
   orf134ng    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIIGG  324
25  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  150
   orf134ng    LVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVACSTGIGIAFGFMPANKAAKLNPIDA  384
30  orf134.pep  LAQD  154
   orf134ng    LAQD  388

```

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACCAT
35 51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
   101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
   151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG
   201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
   251 GCTACGTTGC CTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC
40 301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
   351 TTTGACGCTG CGCGGGCTGA AGCTGGAAC GGGGCGGCTG TTTGATGAGA
   401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
   451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
   501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAAGAC GAAACGCTT
   551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGCTGATG
45 601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
   651 AGACAATGCC AATACCGGG TTGCCGAAAA AGGGCTGGCC GAGCTGCTCA
   701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
   751 AGGCAGATGG TCGAAAGCAC CACCGGTACG ATGAAGTCG TGATTTCCCTC
50 801 CATCGCCCTG ATTTCAATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
   851 TGCTGGTGTG CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
   901 ATCGGCGCGC GCGCGGCAA TATTTTGCAG CAGTTTTTGA TFGAGGCGGT
   951 GTTAATCTGC ATCATCGGAG GCTTGGTCGG CGTAGGTTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTG
1051 GCGGCATCCG TTATCGGGGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
55 1101 GTTCGGCTTT ATGCCTGCCA ATAAGGCAGC CAAACTCAAT CCGATAGATG
   1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 538>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSMGT
60 51  NTISIFPGRG FGDRRSQKIK TLTIDDAKII AKQSYVASAT PMTSSGGTIL
   101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
   151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
   201 HQITGESHTN SITVKIKDNA NTRVAEKGVA ELLKARHGE DFFMNSDSI
   251 RQMVSTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
   301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFPMDIS

```

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5  orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSMGNTNTISIFPGRG
   orf134-1     MSVQAVLAHKMRSLLTMLGIIIGIASVVSVALGNGSQKKILEDISSIGTNTISIFPGRG

   orf134ng      FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
10  orf134-1     FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

   orf134ng      RGLKLETGRLEFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1     RGLKLETGRLEFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

15  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT
   orf134-1     ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGT

20  orf134ng      DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
   orf134-1     DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

   orf134ng      IGARRGNILQQFLIEAVLICIGGLVGVLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
25  orf134-1     IGARRGNILQQFLIEAVLICIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC

   orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
   orf134-1     STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

sp|P75831|YBJZ ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length = 648

```

35  Score = 297 bits (753), Expect = 6e-80
   Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRG 60
M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
Sbjct: 260 MAWRALAANKMRTLLTMLGIIIGIASVVSIVVVGDAAKQMVLAIRSIGTNTIDVYPGKD 319

40  Query: 61 FGDRRSGKIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
      + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
Sbjct: 320 FGDDDPQYQALKYDDLIAIQKPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRLEFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTILFRKRPLTVIGVMKK 179
      G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
Sbjct: 380 YGMTFSEGNTFNQQLNGRAQVVVLDNTRRQLFPHKADVGEVILVGNMPARVIGVAEE 439

50  Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
      ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
Sbjct: 440 KQSMFGSSKVLRVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSEAEQQLTRLLSLRHGK 499

55  Query: 240 EDFFMNNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRM 299
      +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EIGIRM
Sbjct: 500 KDFFTWNMDGVLTVEKTTTTLQLFLTLVAISLVVGGIGVMNIMLVSVTERTREIGIRM 559

60  Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
      A+GAR ++LQQFLIE F+ + + S +++ A
Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAF 619

Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
CST GI FG++PA AA+L+P+DALA++
Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1  ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
51  GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101 TTTCCCTTCCT GATTTTGAAG GAACGGATT TCCGTTACAC GCAGGCGGTG
151 CTGCTCCTTG GTTTTGCCCG CGTGGTATTG CTGCTTAATC CCTCGTTCGG
10  201 CAGCGGTGAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
301 GGCTGGCGCG TCGTGTTTTA CCTTCCGTG ACAGGTGTGG CGATGTCGTC
351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
401 TTTATCTGTC GTGCATCGGC GTGTCCGCGC TGATTGCCCA ACTGTCGATG
15  451 ACGCGCGCCT ACAGAGTCGG CGACAAATC ACGGTGCGCT CGCTTTCCTA
501 TATGACCGTC GTTTTTCCTG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
601 ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

20  1  ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV
51  LLLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
101 GWRVVFYLSV TGVMSSVWA TLTGWHTLSF PSAVYLSGIG VSALIAQLSM
151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELFWQEI LGMCIISAV
201 F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GGCGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTGTCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA
151 ACCGTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
30  201 GCCCATTGG AAAAACCACT TAAACGCGAG TATGGTCGGG ACGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTATTG GCGGTATTTT CCTTCTGAT
351 TTTGAAAGAA CGGATTTCCG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTTCCGCGC CGGTGAGGAA
35  451 ACGGCGGCAC TCGCCGGGCT GCGGCGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAATGT CTTTGGCGGG CGAACC CGGC TGCGCGTCCG
551 TGTTTTACCT TTCCGTGACA GGTGTGCGCA TGTCGTCCGT TTGGGCGACG
601 CTGACCGGCT GGCACACCT GTCCCTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
40  701 AAGTCGGCGA CAAATTCACG GTTGCCCTCG TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCG ATTTTCTCTG GCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMLFS
51  TVALGAAAVL RRDxFRTPHW KNHLNRMVGV TGAMLLLFYA VTHLPLATGV
101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVVL LLNPSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
50  201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTV
251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N. meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
orf135a      STVALGAAAVLRRDTRTPHWKNLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIF
              50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
orf135a      LAVFSFLILKERISVYTOAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
              110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
              170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSAASAFFLGEELFWQEILGMCIIISAVFX
orf135a      TRAYKVGDKFTVASLSYMTVVFSAASAFFLAEELFWQEILGMCIIILSGILSSIRPTAF
              230     240     250     260     270     280

25     orf135a      KQRLQSLFRQRX
              290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
51  GCGGGCCTGC TTTACCATTA TGAACGTATT GATTAAAGAG GCATCGGCAA
101 AATTTGCCCT CCGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTCCGCAC
35  201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCCTG TTTACACGCA GCGCGTGCTG CTCCTTGGTT
401 TTGCCCGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAGGAA
40  451 ACGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCCG GCTGGGCGTA
501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
601 TTGCCCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCTGT
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
45  701 AAGTCGGCGA CAAATTACAG GTTGCCTCGC TTTCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51  TVALGAAAVL RRDTFRTPHW KNHLNRSMVG TGAMLLLFYA VTHLPLATGV
101 TLSTSSIFL AVFSFLILKE RISVYTOAVL LLGFAGVVLL LNPFSFRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSICIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
55  251 FSALSAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60  orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||
orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

-318-

5 orf135a.pep RRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
 orf135-1 RRDXFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE

10 orf135a.pep RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
 orf135-1 RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG

15 orf135a.pep WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT

20 orf135a.pep VASLSYMTVVFSALSAAFFLAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
 orf135-1 VASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20 *N.gonorrhoeae*:

orf135.pep GTGAMLLLFYAVTXLPLATGVTLSTSSIF 30
 orf135ng STVTLGAAAVLRRDTFRTPHWKNHLNRSMVGTGAMLLLFYAVTHLPLTTGVTLSTSSIF 335

25 orf135.pep LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK 90
 orf135ng LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK 395

30 orf135.pep VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 150
 orf135ng VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 455

35 orf135.pep TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAVF 201
 orf135ng TRAYKVGDKFTVASLSYMTVVFSALSAAFFLGEELFWQEILGMCIIISAAF 506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

1 MPSEKAFRRH LRTASFQGLH LHHFHQKVGK CGIIGFGIHI FPTLLPAAQG
 51 ILDIQLGLER IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL
 101 NLGHETDTHL IAQARRFIAD FGNIREPMRRG EAKTFCRCFR FDGIDGIHGD
 151 FRQCCHINRL APGKDCRNGK RDKVFETHRH YNQVCLEKTN CSARKIKFRH
 201 KQAKTHSTS LAARFTIRPS LSQRPEMDTA KKDILGSGWM LVAAACFTVM
 251 NVLKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRRDT FRTPHWKNHL
 301 NRSMVGTGAM LLLFYAVTHL PLTTGVTLSTSSIFLAVFS FLILKERISV
 351 YTQAVLLLG AGVLLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS
 401 LAGEPGWRVV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI
 451 AQLSMTRAYK VGDKFTVASL SYMTVVFSAL SAAFFLGEEL FWQEILGMCII
 501 IISAAF*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50 1 ATGGATACCG CAAAAAAGA CATTTAGGA TCGGGCTGGA TGCTGGTGGC
 51 GGCGGCTGCT TTCACCGTTA TGAACGTATT GATTAAAGAG GCATCGGCAA
 101 AATTGCCCCT CGGCAGCGGC GAATTGCTCT TTTGGCGCAT GCTGTTTCA
 151 ACCGTACGCT TCGGTGCTGC CGCCGTATTG CGGCAGGACA CCTTCCGCAC
 201 GCCCATTTGG AAAAACCACT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
 55 251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGAC AACCGGCGTT
 301 ACCCTGAGTT ACACCTCGTC GATTTTtttg GCGGTATTTT CCTTCCTGAT
 351 TTTGAAAGAA CGGATTTCCT TTTACACGCA GGCGGTGCTG CTCCTTGGTT
 401 TTGCCGCGCT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTACAGGAA
 451 CCGGCGGCAC TCGCCGGGCT GCGGCGGCGC GCGATGTCG GCTGGGCGTA
 60 501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
 551 TGTTTTACCT TTCCGCAACC GCGGTGGCGA TGTCGTCggt ttgggcgacg
 601 Ctgaccggct ggCACaccct GTCCTTTcca tcggcagttt ATCtgtCGGG

-319-

5 651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg cGCGcctaca
 701 aaGTGCGCGA CAAATTCACG GTTGCCCTCGC tttcctaTat gaccgtcGTC
 751 TTTTCCGCC TGTCTGCCGC ATTTTCTCTg ggcgaagagc tttTctggCA
 801 GGAATACTC GGTATGTGCA TCATTatccT CAGCGGCATT TTGAGCAGCA
 851 TCCGCCCCAT TGCCTTCAAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

10 1 MDTAKKDILG SGWMLVAAAC FTVMNVLKE ASAKFALGSG ELVFWRLFS
 51 TVTLGAAAVL RRDFTFPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
 251 FSALSAAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
 301 *

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

orfl35ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVLKEASAKFALGSGELVFWRLFSTVTLGAAAVL
 orfl35-1 MDTAKKDILGSGWMLVAAACFTIMNVLKEASAKFALGSGELVFWRLFSTVALGAAAVL
 20 orfl35ng-1.pep RRDFTFPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE
 orfl35-1 RRDxFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
 25 orfl35ng-1.pep RISVYTQAVLLLGFAVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG
 orfl35-1 RISVYTQAVLLLGFAVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
 orfl35ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT
 30 orfl35-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFSAVYLSGIGVSALIAQLSMT RAYKVGDKFT
 orfl35ng-1.pep VASLSYMTVVFSAALSAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR
 orfl35-1 VASLSYMTVVFSAALSAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTCGTCCTG TTCCCGCAGA TAATCCGAGT
 51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTCCG GCACATCGGA
 101 TGCTCTTCCA GATTTTCGGG ATGTTCTTTT TCTTCATACA CCAGCAATAT
 151 CTGCCCGGGA TCGCCGAAAT CGATTCCCCA TCGCGCATCG TGTTCCGGTGC
 45 201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
 301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTCAGTTCG CcGGTTCAT
 351 TGTTCAGCAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
 401 CACATATGTT CGCAAATTTT GCCGTCTTCG CCGTCTTGA AAAAAGGGAC
 451 TTTGACCATG GCAAATCCA AGGCGGAAAT AATGCGGCGG CGTTCCTCAA
 50 501 AAAGcTCGCG CCAAAAATAT TTGAATGTTT TACGGGCGCG TTCGTCGGCA
 551 CGGTTTACCG GTTCGTCTGC CTGTTCTACA TAATAAATGA CGGAATCGCC
 601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTGCA CCTTACTGCG
 651 GCTTCTgCc kTCGGCATCC GATTCCGATT TGAAAAGTTC mmrwyATTCC
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY
 51 LPGIAEIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPVADVNR

-320-

101 NANAFALFDI GQFAXFIVQH TVNIKTIVKIN IVDPHMFANF AVFAVLEKRD
 151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
 51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 GGATGCTCTT CCAGATTTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TATCTGCCCC GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 TCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCCTGTAT GGTAAAGCCG
 10 CCGTAGGGGA TGCCGTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
 251 CCGTAGGGGA TGCCGTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
 301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGGTCAGT TCGCCGGGTT
 351 CATTGTTCAG CACACCGTAA ATATAAGAC CGTCAAAATA AATATCGTCG
 401 ATCCACATAT GTTCGCAAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
 15 501 AAAAAAGCTC GCGCCAAAAA TATTGAATG TTTTACGGGC GCGTTCGTCG
 551 GCACGGTTTA CCGGTTTCGTC TGCCTGTTCT ACATAATAAA TGACGGAATC
 601 GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
 651 CGGCTTTCTG CCTTCGGCAT CCGATTGCGA TTTGAAAAGT TCCAAATATT
 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

1 MMKRRIAVFV LFPOIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ
 51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAVGDAVA HEHPVADVNN
 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
 151 DFDHGKIQGG NNAAFPKKL APKIFECFTG AFGTVYRFV CLFYIINDGI
 25 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N.meningitidis*:

30		10	20	30	40	50	59
orf136.pep		MMKRRIAVFVLFPOIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS					
orf136a		MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS					
35		10	20	30	40	50	60
orf136.pep	60	70	80	90	100	110	119
orf136a	60	70	80	90	100	110	120
40		70	80	90	100	110	120
orf136.pep	120	130	140	150	160	170	179
orf136a	120	130	140	150	160	170	180
45		130	140	150	160	170	180
orf136.pep	180	190	200	210	220	230	
orf136a	180	190	200	210	220	230	

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCCTGTAT GGTAAAGCCG
 251 CCGTAGGGAA TGCCGTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC

-321-

301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT
 351 CATTGTTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG
 401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGCTCT GGAAAAAAGG
 451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GCGGTTCCCA
 501 AAAAAGCTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCCGG
 551 CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATAATAAAT GACGGAATCG
 601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
 651 CGGCTTTCTG CCTTCGGCAT CCGATTCCGA TTTGAAAGT TCCAAATATT
 701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ
 51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNVA HEHPVADV
 101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMFAN FAXFAVLEKR
 151 ALTMAKSKXX XMRRRSQKSS RQKYLNLVRA RSPARFTGLS ACST**MTES
 201 PIISAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
20	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGIAEIDS					
	orf136-1	MMKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS					
		10	20	30	40	50	60
25	orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNVAHEHPVADVNNANAFALFDIGQFAGFIVQ					
	orf136-1	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQFAGFIVQ					
		70	80	90	100	110	120
30	orf136a.pep	HAINVKT VKI NIVDPHMFAN FAXFAVLEKRALTMAKSKXXXMRRRSQKSSRQKYLNLVRA					
	orf136-1	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG					
		130	140	150	160	170	180
35	orf136a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRVRYLFAPYCGFLPSASDSLKS SKYSE					
	orf136-1	AFVGTVYRFVCLFYIINDGIAH---SAPQVRVRYLFAPYCGFLPSASDSLKS SKYSE					
		190	200	210	220	230	
40							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from

N.gonorrhoeae:

45	orf136.pep	MMKRRIAVFVLPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS	59
	orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS	60
	orf136.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQFAXFIVQ	119
50	orf136ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNNANAFALFDIGQSAGFIVQ	120
	orf136.pep	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
	orf136ng	HTVNIKT VKI NIVDPHMFAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKVFEFTG	180
55	orf136.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSXXXSE	234
	orf136ng	AFAGTVYRFVCLFYIINDGIAHTAPQVRVRYLFAPYRGFLPPASDSLKS SKYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

-322-

```

101  GGATGCTCTT  CCAAATTTTC  GGGATGTTCT  TTTTCTTCAT  ACACCGGCAA
151  TACCTGCCCG  GGATCGCCGA  AATCGATTCC  CCAGGCGGTA  TCGTGTTCGG
201  TACGCTCCTC  TTCCGTCATC  TGTCCGCGCA  TTGCCTGTAC  GGTAAAGCCG
251  CCGTAGGGGA  TGCCGTTGCA  CACGAACATC  CAGTCGCTGA  TGTCGCCAAC
5   301  CGGAACGCAA  ACGCTTTCGC  CTTGTTTCGAC  ATTGGTCAGT  CCGCCGGGTT
351  CATGTTCAG  CACACCGTAA  ATATAAGAC  CGTCAAAATA  AATATCGTCG
401  ATCCACATAT  GTTCGCAAAT  TTCGCCGCTC  TCGCCGCTCT  GGAAAAAAGG
451  GACTTTGACC  ATGGCAAAAT  CCAAGGCGGA  AATAATGCGG  CGGCGTTCCC
10  501  AAAAAAGCTC  GCGCCAAAAG  TATTTGAATG  TTTTACGGGC  GCGTTCGCCG
551  GCACGGTTTA  CCGGTTCGTC  TGCCTGTTCT  ACATAATAAA  TGACGGAATC
601  GCCCATCATA  CTGCTCCTCA  ACGTGTACGG  TATCTGTTTG  CACCTTACCG
651  CCGTTTTCTA  CCTCCGGCAT  CCGATTCGGA  TTTGAAAAGT  TCCAAATATT
701  CGGAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 556>:

```

15   1  MMKRRIAVFV  LLMQKIRILG  QLLPKIVNTV  PAHRMLFQIF  GMFFFFIHRQ
51   YLPGIAEIDS  PGGIVFGTLL  FRHLSAHCLY  GKAAVGDAVA  HEHPVADVAN
101  RNANAFALFD  IGQSAGFIVQ  HTVNIKTVKI  NIVDPHMFAN  FAVFAVLEKR
151  DFDHGKIQQG  NNAAAFPKKL  APKVFECFTG  AFAGTVYRFV  CLFYIINDGI
201  AHHTAPQVRV  YLFAPYRGFL  PPASDSLKLS  SKYSE*

```

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

```

orfl36ng      MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
orfl36-1      MMKRRIAVFVLFPPQIIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
25  orfl36ng      PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVANNANAFALFDIGQSAGFIVQ
orfl36-1      PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVANNANAFALFDIGQFAGFIVQ
30  orfl36ng      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDHKGKIQGGNNAAAFPKKLAPKVFECEFTG
orfl36-1      HTVNIKTVKINIVDPHMFANFAVFAVLEKRDHKGKIQGGNNAAAFPKKLAPKIFECFTG
35  orfl36ng      AFAGTVYRFVCLFYIINDGIAHHTAPQVRVRYLFAPYRGFLPPASDSLKSSKYSEX
orfl36-1      AFVGTVYRFVCLFYIINDGIAHHSAPQVRVRYLFAPYCGFLPSASDSLKSSKYSEX

```

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

```

1   ATGGAATA  TGGTAACGTT  TTCAAAAATC  AGACCGCTTT  TGGCAATCGC
51  CGCCGCCGCG  TTGCTTGCCG  CC.TGCGGAC  GGCGGGAAAT  AATGCTGTCC
101 GCAAGCCGGT  GCAAACCGCC  AAACCGCCG  CAGTGGTCGG  TTTGGCACTC
45  151  GGTGGCGGCG  CATCTAAAGG  ATTTGCCCAT  GTAGGTATTA  TTAAGGTTTT
201  GAAAGAAAAC  GGTATTCCTG  TGAAGGTGGT  TACCGGCACC  TCCGCAGGTT
251  CGATTGTCGG  CAACCTTTTT  GCATCGGGTA  TGTCGCCCGA  CCGCCTCGAA
301  TTGGAAGCCG  AAATTTTAGG  CAAAACCGAT  TTGGTCGATT  TAACCTTGTC
351  CACCAATGGG  TTTATCAAAG  GCGCAAAGCT  GCAAATTAC  ATCAACCGAA
401  AACTCCGCGG  CATGCAGATT  CAGCAGTTTC  CCATCAAAAT  TGCCGCC..

```

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

```

1   MENMVTFSKI  RPLLAIAAAA  LLAAXRTAGN  NAVRKPVQTA  KPAAVVGLAL
51  GGGASKGFAH  VGIIKVLKEN  GIPVKVVTGT  SAGSIVGNLF  ASGMSPDRLE
101 LEAEILGKTD  LVDLTLSNG  FIKGAKLQNY  INRKLGRMQI  QQFPIKFAA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

-323-

1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 5 201 GAAAGAAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATTGTCCG CAGCCTTTTT GCATCGGGTA TGTGCCCCGA CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCACTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 10 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTCAATC AGGGGAATGC
 501 CGGGCAGGCT GTGCGCGCTT CCGCGCCAT TCCAATGTG TTCCAACCCG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCCCGCGC GCAGGGGCGC AATTTCGTGA TTGCGTCGA
 651 TATTTCGCCG CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
 15 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAAARRQGA NFVIAVDISA RPKGNISSQF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPO VLDLGAUVGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep		MENMVTFSKIRPLLAIAAAA	LLAAXRTAGN	NAVRKPVQTA	KPAAVVGLAL	GGGASKGFAH	
35	orf137a	MENMVTFSKIRPLLAIAAAA	LLAACGTAGN	NAARKPVQTA	KPAAVVGLAL	GGGASKGFAH	
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf137.pep	VGIKVLKENGIPVKVVTGT	SAGSIVGNLF	ASGMSPDRLE	LEAEILGKTD	LVDLTLSTNG	
	orf137a	VGIKVLKENGIPVKVVTGT	SAGSIVGSLF	ASGMSPDRLE	LEAEILGKTD	LVDLTLSTNG	
		70	80	90	100	110	120
45	orf137.pep	FIKGA	KLQNYINR	KLGRMQI	QQFPIKFAA		
	orf137a	FIKGE	KLQNYINR	KVGGRRI	QQFPIKFAA	AVATDFETGKAV	AFNQGNAGQAVRASAAIPNV
		130	140	149			
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGCAC GGCGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 201 GAAAGAAAAC GGTATTCTTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 55 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTGCCCCGA CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCACTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTCAATC AAGGGAATGC
 60 501 CGGGCAGGCT GTGCGCGCTT CCGCGCCAT TCCAATGTG TTCCAACCCG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

5
 601 CCCGTCAGTG CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA
 651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTG GATCAGAAAA AACGCGCCAT CCGGTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCTGTAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10
 1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRLE
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARAALPE IKRKLAAYRY
 15 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20
 orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
 orf137a.pep VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
 orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTLSTSG
 25
 orf137a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 30
 orf137a.pep FQPVIIGRHTYVDGGLSQPVVPVSAARRXXXXXVIAVDISARPSKNISQGFPSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVPVSAARRQGANFVIAVDISARPGKNISQGFPSYLDQTLNV
 35
 orf137a.pep MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY
 orf137-1 MSVSALQNELGQADVVIKPOVLDLGA VGGFDQKKRAIRLGEEAARAALPEIKRKLAAYRY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40
 orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH 60
 orf137ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60
 45
 orf137.pep VGIIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLELEAEILGKTDLVDLTLSTNG 120
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLELEAEILGKTDLVDLTLSTSG 120
 orf137.pep FIKGAKLQNYINRKLGRMQIQFPIKFAA 149
 50
 orf137ng FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55
 1 ATGGAATA TGGTAACGTT TTCAAAATC AGATCATTTT TGGCAATCGC
 51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GCGGGGAAAC AATGCCGCC
 101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCGC TTGGCACTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTT
 201 GAAAGAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATAGTCGG CAGCCTTTTG GCATCGGTA TGTCGCCCGA CCGCCTCGAA
 301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTTGTC
 351 CACCAAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 60
 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTT CCATCAAATT TGCCGCCGTT
 451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCATC AAGGGAATGC

-325-

501 CGGGCAGGCG GTTCGTGCTT CCGCCGCCAT TCCCAATGTG TTCCAGCCAG
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGGCG AATTTCGTGA TTGCCGTCGA
 651 TATTTCGCA CGTCCGAGCA AAAATGTCCG TCAAGGTTTC TTCTCTTATC
 5 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTGCA AAACGAGTTG
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gttTTGGATT TGGGTGCAGT
 801 CCGCGGATTC GATCAGAAAA AGCGCGCCAT CCGGTGGGC GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSLL ASGMSFDRLE
 101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAQQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV
 15 201 PVSAARRQGA NFVIAVDISA RPSKNVGQGF FSYLDQTLNV MSVSVLQNEL
 251 GQADVVIKQ VLDLGAVGGF DQKKRAIRLG EEAAARAALPE IKRKLAAARY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAAALLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH
 orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSFDRLELEAEILGKTDLVDLTSTSG
 25 orf137-1 VGIIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSFDRLELEAEILGKTDLVDLTSTSG
 orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAQAVRASAAIPNV
 30 orf137ng FQPVIIGRHKYVDGGLSQPVPVSAARRQGANFVIAVDISARPSKNVGQGF FSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVPVSAARRQGANFVIAVDISARPGKNISQGF FSYLDQTLNV
 35 orf137ng MSVSVLQNELGQADVVIKQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
 orf137 MSVSALQNELGQADVVIKQVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site
 (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and
 40 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for
 raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
 51 CATCCTGTTG ACCGCCCTGC TCAATGCCT CTCCTGcTG CCGCTTTCCT
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTCGCMAT ATGCGGCAGG CGGGTTTGAA
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TGCGGAAACG GCAAAGGCG
 251 GTTTGGAAGT TGCCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
 50 301 ATGTTCAAAG CGGTACACG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
 351 ACACGAAGGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCHLTLGN RLGLAFYLL
 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDTET
 55 101 MFKAHVHGEH VQQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
  101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
  151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGGCAGG  CGGGTTTGAA
  201  CCCCACCCC   AAAACGGTCA  AAGCCGTTT  TCGGGAACG  GCAAAAGGCG
  251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCAGGAAG  CATAGAAACA
  301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGACAA
  351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGG
  401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
  451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
  501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
  551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCACCAC
  601  GTCCCTCCC   CTCAAGAAGG  CGGGGAAGGC  GTATGGGTGG  ATTTCTTCGG
  651  CAAACCTGCC  TATACCATGA  CGCTGGCGGC  AAAATTGGCA  CACGTCAAAG
  701  GCGTGAAAC   CCTGTTTTTC  TGCTGCGAAC  GCCTGCCTGG  CGGACAAGGT
  751  TTCGATTGTC  ACATCCGCCC  CGTCCAAGGG  GAATTGAACG  GCGACAAAGC
  801  CCATGATGCC  GCCGTGTTCA  ACCGCAATGC  CGAATATTGG  ATACGCGGTT
  851  TTCCGACGCA  GTATCTGTTT  ATGTACAACC  GCTACAAAAT  GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1  MFRLOFRLFP  PLRTAMHILL  TALLKCLSL  LPLSCLHTL  GNRLGHLAF  YLLKEDRAR  IVAX
    51  KEDRARIVAN  MRQAGLNPD  PDKTVKAV  FAETAKGG  LEAPAFRKP  EDIETMFK  AVHGWEHV  QQALDKHE  G
  101  MFKAVHGWEH  VQALDKHEG  LLFITPHIG  SYDLGGRI  SQQLPFPL  TAMYKPKI  KAIKIMQ  AGRVRGK  GKTAPT  SIQ
  151  KPPKIKAIK  IMQAGRVRG  KGTAPT  SIQVKQI  IKALRS  GEATIV  LPPDH
  201  VPSQEGGEG  VWVDFGK  PAVTMT  LAAKLH  VKGK  VTLFF  CCERL  PGGQG
  251  FDLH  IRPVQG  ELNGD  KAHDA  AVFN  RNAEY  WIRRF  PTQY  LFMYN  RYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep  MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
orfl38a     MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCLHTLGNRLGHLAFYLLKEDRARIVAX
      10      20      30      40      50      60
      70      80      90     100     110     120
orfl38.pep  MRQAGLNPDPKTVKAVFAETAKGGLELAPAFRKPEDIETMFKAVHGWEHVQQALDKHEG
orfl38a     MRQAGLNPDPKTVKAVFAETAKGGLELAPAFRKPEDIETMFKAVHGWEHVQQALDKHEG
      70      80      90     100     110     120
orfl38.pep  LLF
orfl38a     LLFITPHIGSYDLGGRIYQQLPFPLTAMYKPKIKAIKIMQAGRVRGKGTAPT  SIQ
      130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1  ATGTTTCGTT  TACAATTCAG  GCTGTTTCCC  CCTTTGCGAA  CCGCCATGCA
    51  CATCCTGTTG  ACCGCCCTGC  TCAAATGCCT  CTCCCTGCTG  CCGCTTTCCT
  101  GTCTGCACAC  GCTGGGAAAC  CGGCTCGGAC  ATCTGGCGTT  TTACCTTTTA
  151  AAGGAAGACC  GCGCGCGCAT  CGTCGCCAAT  ATGCGTCAGG  CAGGCATGAA
  201  TCCCACCCC   AAAACGGTCA  AAGCCGTTT  TCGGGAACG  GCAAAAGGCG
  251  GTTTGGAAct  TGCCCCCGCG  TTTTTCAGAA  AACCAGGAAG  CATAGAAACA
  301  ATGTTCAAAG  CGGTACACGG  CTGGGAACAT  GTGCAGCAGG  CTTTGACAA
  351  ACACGAAGGG  CTGCTATTCA  TCACGCCGCA  CATCGGCAGC  TACGATTGG
  401  GCGGACGCTA  CATCAGCCAG  CAGCTTCCGT  TCCCGCTGAC  CGCCATGTAC
  451  AAACCGCCGA  AAATCAAAGC  GATAGACAAA  ATCATGCAGG  CGGGCAGGGT
  501  TCGCGGCAAA  GGAAAAACCG  CGCCTACCAG  CATACAAGGG  GTCAAACAAA
  551  TCATCAAAGC  CCTGCGTTCG  GGCGAAGCAA  CCATCGTCCT  GCCCACCAC

```

-327-

601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
 701 GCGTGAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
 751 TTCGATTGTC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
 5 801 CCATGATGCC GCCGTGTTC ACCGCAATGC CGAATATTGG ATACGCCGTT
 851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LPLSCLHTLGN RLGHIAFYLL
 51 KEDRARIVAN MRQAGLNPD P KTVKAVFAET AKGGLLELAPA FFRKPEDIET
 10 101 MFKAVHGWEH VQQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY
 151 KPPKIKAIK IMQAGRVRGK GKTAPT SIQ VQI I KALRS GEATIVLPDH
 201 VPSPQEGGEG VWVDFFGKPA YMTLAAKLA HVKGVKTLFF CCERLPGGQG
 251 FDLHIRPVQG ELNGDKAHD A VFNRNAEYW IRRFPTQYLF MYNRYKMP*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHIAFYLLKEDRARIVAN
 orf138-1 MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHIAFYLLKEDRARIVAN
 20 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
 orf138-1 MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG
 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPT SIQ
 25 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPT SIQ
 orf138a.pep VKQI I KALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
 orf138-1 VKQI I KALRSGEATIVLPDHVPSPQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGVKTLFF
 30 orf138a.pep CCERLPGGQGF DLHIRPVQGE L NGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP
 orf138-1 CCERLPGGQGF DLHIRPVQGE L NGDKAHDAAVFNRNAEYWIRRFPTQYLFMYNRYKMP

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

orf138.pep MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHIAFYLLKEDRARIVAX 60
 40 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSL LPLSCLHTLGNRLGHIAFYLLKEDRARIVAN 60
 orf138.pep MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG 120
 orf138ng MRQAGLNPD P KTVKAVFAETAKGGLLELAPAFFRKPEDIETMFKAVHGWEHVQQALDKHEG 120
 45 orf138.pep LLF 123
 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQ 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
 51 CATCCTGTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT
 101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATCGGCGAGG CGGGTTTGAA
 201 CCCCACACG CAGACGGTCA AAGCCGTTT TGCGGAAACG GCAAAATGCG
 55 251 GTTTGGAAC TGCCTCCGCG TTTTCAAAA AACCGGAAGA CATCGAAACA
 301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
 351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGTTG
 401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCCATGTAC
 451 AAGCCGCCGA AAATCAAGC GATAGACAAA ATCATGCAGG CGGGCAGGTT
 60 501 GCGCGGCAAA GGCAAAACcg cggccaccgg catACAAGGG GTCAAACAAA
 551 tcatcaAGGC CCTGCGCGCG GCGAGGCAA CCATcATCCT GCGCGACCA

-328-

5 601 GTCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA
 651 ACCTGCATAC acCATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG
 701 TGAAGAACCT GTTTTCTGC TGCAGACGCC TGCCCGACGG ACAAGGCTTC
 751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAGCCCA
 801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC
 851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10 1 MFRLQFRLFP PLRTAMHILL TALLKCLSL SLSCSLHTLGN RLGHAFYLL
 51 KEDRARIVAN MRQAGLNPDQT VTKAVFAET AKCGLELAPA FFKKPEDIET
 101 MFKAVHGWEH VQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY
 151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQ VKQIIKALRA GEATIILPDH
 201 VPSQEGGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDQGF
 251 VLHIRPVQGE LNGNKAHDAA VFNRNTEYWI RRFPTQYLFM YNRYKTP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSLHTLGNRLGHAFYLLKEDRARIVAN
 orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSLSLSCSLHTLGNRLGHAFYLLKEDRARIVAN
 20 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGWEHVQALDKHEG
 orf138ng MRQAGLNPDQT VTKAVFAETAKCGLELAPAFFKKPEDIETMFKAVHGWEHVQALDKGEG
 25 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
 orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG
 30 orf138-1.pep VKQIIKALRSGEATIVLPDHVPSQEGGEGVWVDFFGKPAYTMTLAAKLAHVKGKTLFF
 orf138ng VKQIIKALRAGEATIILPDHVPSQEGG-GVWADFFGKPAYTMTLAAKLAHVKGKTLFF
 35 orf138-1.pep CCERLPGGQGFDLHIRPVQGEGLNGDKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKMP
 orf138ng CCERLPDQGFVLHIRPVQGEGLNGNKAHDAAVFNRNTEYWIRRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
 Score = 80.8 bits (196), Expect = 9e-15
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
 40 Query: 101 MFKAVHGWEHVQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159
 + + V G E +++AL G+G++ IT H+G+++ L Y SQ P Y+PPK+KA+D
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLAVD 150
 Query: 160 KIMQAGRVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHVPSQEGGGVWADFFGKPA 219
 ++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A
 45 Sbjct: 151 ELLRKQRVQLGNKVAASTKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVFFATQA 208
 Query: 220 YMTLAAKLAHVKGKTLFFCCERLPDQGF 250
 T + +F RLPDG G+
 Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A
 55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5      1  ..GCGTGGTCCG CCGGCGAATC GTGGCGTGTG TTAATGGAAG GTGAAACGTG
      51  GCATGCGGTG TGGAATACTT TGCCTTCTC GCGCGCGCG GTGTATGCGG
     101  CAGCGGTTTT GGGTGTGGTG TATGCGGCGC GCGCGCGCG GTGCGGCTGG
     151  ATGCGCGGGG TGATGTTTTA GCCGTTATG GTGTGCGCGG TTTGTGTTTC
     201  GCGCGGCGTG CTGCTGCTTT ATCCGCAGTG GACGGCTTCG TTGCCGTTGC
    10  251  TGCTGGCGAT GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
     301  TTATCAGCCT GGGATGCACT GCCGCGGAT TACGGCAGGG CCGCGCGCGG
     351  TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
     401  TGAAACCGGC GTTGCAGCGC GGTCTGACTT TGGCGCGGCG AACCTGCGTG
     451  GCGCAATTTG CCGCGACATT GTTCTGTGCG CGTCCGGAAT GGCAGACGCT
    15  501  GACGACTTTG ATTTATGCTT ATTTGGGACG CGCGGGTGAG GATAATTACG
     551  CGCGGGCGAT GGTGCTG..
  
```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAG VYAAAVLGVV YAAPARRSAW
     51  MRGLMFXPFM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
    101  LSAWDALPPD YGRAAAGLGA NGFQTACRIT FPLLLPALRR GLTLAAATCV
    151  GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

      1  ATGGATGGAC GCGCTTGGGT GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
     51  GGCTTTTGTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
    101  ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCTTA TATGCTCAAA
     151  CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
     201  GCCTTTGGGC GTGCCGTGTC CGTGGGTGCT GCGCGGCGTG GCGTTTCCGG
     251  GCGGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCCACG
     301  TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
    30  351  GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
     401  TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
     451  GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CCGGGGCGTG
     501  GCGGGCGGTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
     551  GCGGCGGTGT CCTTGTCTTT CTGTATGTT TTTCCGGGTT CCGGCTGCGC
    35  601  CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
     651  GTTGGTCATG TTCGAACTCG ATATGGCGGT TGCTTCGGTG CTGGTGTTGG
     701  TGGTGTGGG GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
     751  AGGCGCGCGG TTTCCGATAA GCGGTTTCC CTTGTGATGC CGTCGCGGCC
     801  GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTCGCGCG GCGGTGTTGT
    40  851  CTGTGTGCTG CCTGTTTCCT TTGTTGCAA TTGTTGTGAA AGCGTGGTCG
     901  GCCGGCGAAT CGTGGCGTGT GTTAATGGAA AGTGAACCGT GGCAGGCGGT
     951  GTGGAATACT TTGCGCTTCT CCGCGGCGGC GGTGTATGCC GCGGCGGTTT
    1001  TGGGTGTGTT GTATGCGGCG GCGGCGCGGC GGTCCGCGTG GATGCGCGGG
    1051  CTGATGTTTT TGCCGTTTAT GGTGTCGCGG GTTGTGTTT CCGCGGGCGT
    45  1101  GCTGCTGCTT TATCCGAGT GGACGGCTTC GTTGCCGTTG CTGCTGGCGA
     1151  TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
     1201  TGGGATGCAC TGCCGCGCGA TTACGGCAGG GCGGCGCGCG GTTGGGTGTC
     1251  AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCTC TTGAAACCGG
     1301  CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGCAGAAATT
    50  1351  GCGGCGACAT TGTTTCTGTC GCGTCCGGAA TGGCAGACGC TGACGACTTT
     1401  GATTTATGCC TATTTGGGAC GCGCGGTTGA GGATAATTAC GCGCGGGCGA
     1451  TGGTGTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT TTTCTGCTG
     1501  TTGGACGGCG GCGAAGCGG AAAACAGACG GAAACGTTAT AA
  
```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRRWVWVG AFALLPSAFL AVMVVAPLWA VAAYDGLAWR AVLSDAYMLK
     51  RLAWTVFQAA ATCVLVPLG VPVAWLARL AFPGRALVLR LLMLPFVMP
    101  LVAGVGVLLAL FGADGLLWRG RQDTPYLLLY GNVFFNLPLV VRAAYQGFVQ
    151  VPAARLQATAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
  
```

Computer analysis of this amino acid sequence gave the following results:

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

	1	ATGGATGGAC	GGCGTTGGGC	GGTATGGGGT	GCTTTTGCCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGGCAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGCGGT
	101	ATGACGGTTT	GGCGTGGCGC	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
40	151	CGTTTGGCGT	GGACGTATT	TCAGGCAGCG	GCAACTGTG	TGCTGGTGCT
	201	GCCTTTGGCG	GTGCCGTGCG	CGTGGGTGCT	GGCGCGGCTG	GCGTTTCCGG
	251	GGCGGGCTTT	GGTGTGCGC	CTGCTGATGC	TGCTTTTGT	GATGCCACG
	301	TTGTGGCGGG	GCGTGGGCGT	GCTGGCTCTG	TTCGGGCGGG	ACGGCCTGTN
	351	GTGGCGCGGC	TGGCAGGATA	CGCCGTATCT	GTTGTGTAC	GGCAATGTGT
45	401	TTTTTNACCT	TCCTTGTTTG	GTCAGGGCGG	CATATCAGGG	GTTTGTGCAA
	451	GTGCTGCGG	CACGGCTTCA	GACGGCACNG	ACATTGGGCG	CGGGGGCGGT
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGGCCCGG	TGGCTTGCCG
	551	GCGGCGGTGT	CCTTGCTTTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
	601	TTGCTGCTGG	GCGGCGCGCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
50	651	TTTGCTATG	TTCGAACTCG	ATATGGCGGT	TGCTTCGGTG	CTNGTGTGGC
	701	TGGTGTNNGG	GGTAACNCGG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GGCNGTTTCC	CCTGTGATGC	CGTCGCGGCC
	801	GCAGTCGGTG	GGGGAATATG	TGCTNCTGGC	GTTTGCGGCG	CGGTTGTNGT
	851	CTGTGTGCTG	CCGTGTTTCT	TTGTTGGCAA	TTGTTGTGAA	AGCGTGGCTG
55	901	GCCGCGCAAT	CGTGGCGTGT	GTTAATGGAA	AGTGAACAGT	GGCAGGCGGT
	951	GTGGAATACT	NTGCGCTTCT	CGGCGGCGGC	GGTGTATGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGCG	GGTGCGCGTG	GATGCGCGGG
	1051	CTGATGTTTT	TGCCGTTTAT	GGTGTGCGCG	GTTTGTGTTT	CGGCGGGCGT
	1101	GCTGCTGCTT	NATCCGCAGT	GGACGGCTTC	GTTGCCGCTG	CTGCTGGCGA
60	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAGATGT	TTTATCAGCC
	1201	TGNGATGAC	TGCCGCGCGA	TTACGGCAGG	GCGGCGGCGG	GTTTGGGTGC
	1251	AAACGCTTT	CAGACGGCAT	GCCGCATCAC	GTTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	CGGTCTGACT	TTGGCGGCGG	CAACCTGCGT	GGGCGAATTT
	1351	GCGGCAACCT	TGTTCTNTGT	GCGTCNCGAG	TAGGCGACGC	TGACGACTTT

5	1	MDGRRWAVWG	AFALLPSAFL	AAMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFOAA	ATCVLVLP LG	VPVAWVLARL	AFPGRALVLR	LLMLPFVMPT
	101	LVAGVGV LAL	FGADGLXWRG	WQDTPYLLLY	GNVGFKL PVL	VRAAYQGFVQ
	151	VPAARLQTAX	TLGAGAWRRF	WDIEMPVLRP	WLAGGCVL VF	LYCFSGFGLA
	201	LLLGGSRYAT	VEVEIYQLVM	FELDMAVASV	LVWLXVGUTA	AAGLLYAWFG
10	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFAA	AVXSVCCLFX	LAIATVVKAVS
	301	AGESWRVLME	SETWQAVVNT	XRFSAAAVYA	AAVLGVVYAA	AARRSAWMRG
	351	LMFLPFMVSP	VCVSAGVLLL	XPQWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	XDALPPDYGR	AAAGLGANGF	QTACRTTFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFXSRXE	WQTLTTLIYA	YXGRAGXDNY	ARAMVLTL LL	AAFALGXFLL
15	501	LDGGEKKRT	ETL*			

[illegible]

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

orf139.pep	AWSAGESWRVLMSESETWHAVWNTLRFSAAA	30
	:	
orf139ng	QSVGEYVLLAFSVAVLVSVCCFLPFLSAIVVKAWSAGESRRVLMSESETWQAVWNTLRFSAAA	327
orf139.pep	VYAAAVLGVVYAAPARRSAWMRGLMFXPFMVSPVCVSAGVLLLYPQWTASLP LLAMAYAL	90
	I : : : : : : : : : :	
orf139ng	VFAAAVLGVVYAAAARRLVMRGLVELPFMVSPVCVSAGVLLLYPCGWTASLP LLAMAYAL	387

orf139.pep LAYPFVAKDVL~~SAWDALPPDY~~GRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV 150
 orf139ng LAYPFVAKDVL~~SAWDALPPDY~~GRAAAGLGANGFQTACRITFPLLKPALRRGLTLAAATCV 447

5 orf139.pep GEFAATLFLSRPEWQTLTTLLIYAYLGRAGEDNYARAMVL 189
 orf139ng GEFAATLFLSRPEWQTLTTLLIYAYLGRAGEDNYARAMVL~~LLLLSAFAVCIFLLLDN~~EGG 507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10 1 MDGRCAVVRG AFSLLPSAFL AVMVVA~~PLWA~~ VAAYDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVPLG VPAWVLARL AFGPRALVLR LLMLPFVMP
 101 LVAGVGV~~LAL~~ FGADGLLWRG RQDTPYLLLY GNVFFNL~~PVL~~ VRAAYQGF~~AQ~~
 151 VPAARLQTAR TLGAGAWRPF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
 201 LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
 15 251 RRAVSDKAVS PVMPSPQSV GEYVLLAFSV AVLSVCCLFP LSAIVVKAWS
 301 AGESRRVLME SETWQAVWNT LRFSA~~AAVFA~~ AAVLG~~VVYAA~~ AARRLVWMRG
 351 LVFLPFMVSP VCVSAGVLLL YPGWTASLPL LLAMYALLAY PFVAKDVL~~SA~~
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
 451 AATLFLSRPE WQTLTTLLIYA YLGRAGEDNY ARAMVL~~LLLL~~ SAFAVCIFLL
 20 501 LDNGEGGKRT ETL*

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

1 ATGGATGGAC GGTGTTGGGC GGTACGGGGT GCTTTTTCCC TGCTGCCTTC
 51 GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
 25 101 ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAA
 151 CGTTTTGCGT GGACGGTGTT TCAGGCGGCG GCAACCTGTG TGCTGGTGCT
 201 GCCTTTGGGC GTGCCTGTCTG CGTGGGTGCT GGCGCGGCTG CCGTCCCGG
 251 GGCGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCGTGTGT GATGCCACG
 301 CTGGTGGCGG GCGTGGGCGT GCTGGCTCTG TTCGGGGCGG ACGGGCTGTT
 351 GTGGCGCGGC CGGCAGGATA CGCCGTATCT GTTGTGTGAC GGCAATGTGT
 30 401 TTTTCAACCT GCCCGTGTG GTCAGGCGCG CGTATCAGGG GTTGTCTCAA
 451 GTGCCTGCGG CACGGCTTCA GACGGCACGG ACGTTGGGCG CGGGGGCGTG
 501 GCGGCGGTTT TGGGACATTG AAATGCCCGT TTTGCGCCCG TGGCTTGCCG
 551 GCGGCGGTGT CTTGTCTTTC CTGTATTGTT TTTGCGGGTT CGGGCTGGCA
 601 TTGCTGTTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
 35 651 GTTGGTTATG TTCGA~~ACTCG~~ ATATGGCGGG GGCTTCGGCG CTGGTGTGGC
 701 TGGTGTGGG GGTAA~~CGGCG~~ GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
 751 AGGCGCGCGG TTTCGGATAA GGCGGTTTCC CCCGTGATGC CGTCGCCGCC
 801 GCAATCGGTG GGGGAATATG TATTGCTGGC ATTTTCGGTG GCGGTGTGTG
 851 CCGTGTGCTG CCTGTTTCCT TTGTGCGCAA TTGTTGTGAA AGCGTGGTCG
 901 GCCGGCGAAT CGCGGCGGTG GTTAATGGAA AGTGAACGT GGCAGGCAGT
 951 TGGGAATACt ttGCGCTTTT CGGCGGCGGC GGTGTTTGCG GCGGCGGTTT
 1001 TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGCTGGTGTG GATGCGCGGA
 1051 CTGGTGTGTT TACCGTTTAT GGTGTCGCCG GTTGTGTGTT CGGCGGGCGT
 1101 GCTGCTGCTT TATCCGGGGT GGACGGCTTC GTTACCGCTG CTGCTGGCGA
 45 1151 TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCGGCC
 1201 TGGGATGCAC TGCCGCGCGA TTACGGCAGG GCGGCGGCAG GTTGGGCGC
 1251 AAACGGCTTT CAGACGGCAT GCCGTATCAC GTTCCCCTC TTGAAACCGG
 1301 CGTTGCGGCG CCGTCTGACT TTGGCGGCGG CGACGTGTGT GGGCGAATTT
 1351 GCGGCAACCT TGTTCCTGTC GCGTCCGGAA TGGCAGACGT TGACGACTTT
 50 1401 GATTTATGCC TATTTGGGGC GTGCGGGTGA GGACAATTAT GCGCGGGCAA
 1451 TGGTGTGTGAC ATTGCTGTTG TCGGCATTTG CCGTGTGCAT TTTCTGCTG
 1501 TTGGACAACG GCGAAGGCGg aaaACGGACG GAAACGTTAT AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55 1 MDGRCAVVRG AFSLLPSAFL AVMVVA~~PLWA~~ VAAYDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVPLG VPAWVLARL AFGPRALVLR LLMLPFVMP
 101 LVAGVGV~~LAL~~ FGADGLLWRG RQDTPYLLLY GNVFFNL~~PVL~~ VRAAYQGF~~AQ~~
 151 VPAARLQTAR TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
 201 LLLGGSRYAT VEVEIYQLVM FELDMAGASA LVWLVLGVTA AAGLLYAWFG
 60 251 RRAVSDKAVS PVMPSPQSV GEYVLLAFSV AVLSVCCLFP LSAIVVKAWS
 301 AGESRRVLME SETWQAVWNT LRFSA~~AAVFA~~ AAVLG~~VVYAA~~ AARRLVWMRG
 351 LVFLPFMVSP VCVSAGVLLL YPGWTASLPL LLAMYALLAY PFVAKDVL~~SA~~
 401 WDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
 451 AATLFLSRPE WQTLTTLLIYA YLGRAGEDNY ARAMVL~~LLLL~~ SAFAVCIFLL
 501 LDNGEGGKRT ETL*

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5  orf139ng      MDGRCWAVRGAFSLLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
   orf139-1     MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
10  orf139ng      ATCVLVLPVGVPVAVVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
   orf139-1     ATCVLVLPVGVPVAVVLARLAFPGRALVLRLLMLPFVMPTLVAGVGVLALFGADGLLWRG
15  orf139ng      RQDTPYLLLYGNVFFNLPLVLVRAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
   orf139-1     RQDTPYLLLYGNVFFNLPLVLVRAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
20  orf139ng      WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAGASALVWLVLGVTA
   orf139-1     WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAGASVWLVLVLGVTA
25  orf139ng      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLSAIVVKAWS
   orf139-1     AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCLFPLLAIVVKAWS
30  orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
   orf139       AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
35  orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
   orf139-1     VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAAGLGANGF
40  orf139ng      QTACRITFPLLPALRRGLTLAAATCVGEFAATFLSRPEWQTLTTLIYAYLGRAGEDNY
   orf139-1     QTACRITFPLLPALRRGLTLAAATCVGEFAATFLSRPEWQTLTTLIYAYLGRAGEDNY
45  orf139ng      ARAMVLTLLLSAFAVCIFLLLDNGEGGKRTETL
   orf139-1     ARAMVLTLLLSAFALGIFLLLDGEGGKQTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45  1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
   51  GCGCGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAGA  TTCCGCATCC
   101 ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTGGCAACCC
   151 GGTTTGCCCA  CAGGCAGCAT  TGTCAAAGAC  ATACTGGTCA  AAAACTTCGG
   201 CGGCACGCTC  GCGGCGCTGG  CGCTTCTGGT  CGGCCTGGGC  GCGATGCTCG
   251 AACGTTTGGT  C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50  1  MDGWTQTLTA  QTLGISAAA  IILILILIVR  FRIHALTLV  IVSLLTALAT
   51  GLPTGSIVKD  ILVKNFGGTL  GGVALLVGLG  AMLERLV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55  1  ATGGACGGCT  GGACACAGAC  GCTGTCCGCG  CAAACCCTGT  TGGGCATTTC
   51  GCGCGCGGCA  ATCATCCTCA  TTCTGATTTT  AATCGTCAAA  TTCCGCATCC
   101 ACGCGCTGCT  GACACTGGTC  ATCGTCAGCC  TGCTGACGGC  TTGGCAACCC
   151 GGTTTGCCCA  CAGGCAGCAT  TGTCACGAC  ATACTGGTCA  AAAACTTCGG

```

-334-

201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTCTTTCGA TGCCGGACTA ATCGTCATGC
 5 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTCGCC
 501 GCCCCATCCG GCGCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
 10 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TGCGGACGAA ACCTGGGTTT AGACGCGAAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCGTA TTGGTCGCAC TGTTTGCTT GGGACGCAAA
 15 901 CGCGCGGAAA GCGGCAGCGC GTTGGAAGAA ACCGTGGACG GCGCACTCGC
 951 CCCCCTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC
 1001 GCGTTTTCGG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTCG CCTTGGCACT
 1101 CGGTATCGCG CAAGGTTCGG CAACCTCGC CCTGACCAC GCCGCCGCGC
 20 1151 TGATGGCTCC TGCCGTGTC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTCG GTCGGTTCG GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLISA QTLGISAAAA IILILILIVK FRIHALTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 30 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 251 IFLNTGVSAI ISEKLVSADL TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESGSALEK TVDGLAPVC SVILITGAGG MFGGVLRASG IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGCSEFNDG FWLVGRLLDM DVPTTLKWT VNQTLIALIG
 35 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

meningitidis:

40 orf140.pep 10 20 30 40 50 60
 orf140a MDGWTQTLISAQTLGISAAAAIILILILIVKFRHALTLVIVSLLTALATGLPTGSIVND
 45 10 20 30 40 50 60
 orf140.pep 70 80
 orf140a ILVKNFGGTLGGVALLVGLGAMLERLV
 50 70 80 90 100 110 120
 orf140a VLVKNFGGTLGGVALLVGLGAMLRIVETSGGAQSLADALIRMFGEKRAPFALGVASLIF

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
 51 GCGGCGGCGA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
 55 151 GGTTTGCCCA CAGGCAGCAT TGTCAACGAC GTACTGGTCA AAAACTTCGG
 201 CGGCACGCTC GCGGCGGTGG CGCTTCTGGT CCGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGGCGGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTCTTTCGA TGCCGGACTA ATCGTCATGC
 60 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTCGCC

5 501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAACTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTCTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TCGGACGCAA ACCTGGGTTT AGACGGCAAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA
 901 CGCGGCGAAA GCGGCAGCGC GTTGAAAAA ACCGTGGACG GCGCACTCGC
 10 951 CCCCGTCTGT TCCGTGATTG TGATTACCGG CGCGGGCGGT ATGTTCCGGC
 1001 GCGTTTTCGG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
 1051 GATTGGGCA TTCCCGTCCT TTTGGGCTGT TTCCTGTGCG CCTTGGCACT
 1101 GCGTATCGCG CAAGGTTTCG CAACCGTCGC CCTGACCACC GCCGCCGCGC
 1151 TGATGGCTCC TGCCGTGCGC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
 15 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCGGTTGCA GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGCCT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATCGGC
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLSA QTLIGISAAA IILILILIVK FRIHALLTLV IVSLLTALAT
 51 GLPTGSIVND VLVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMKQDVLP
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 25 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGECSALEK TVDGALAPVC SVILITGAGG MFGGVLRASG IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGCSHFNDG FWLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLFAI V*

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orfl40-1.pep MDGWTQTLSAQTLIGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
 orfl40a MDGWTQTLSAQTLIGISAAAIIILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND 60
 35 orfl40-1.pep ILVKNFGGTLGGVALLVGLGAMLGRLVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF 120
 orfl40a VLVKNFGGTLGGVALLVGLGAMLGRLVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF 120
 40 orfl40-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 180
 orfl40a GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG 810
 orfl40-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV 240
 45 orfl40a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNDLPKEPAKAGTV 240
 orfl40-1.pep VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
 orfl40a VAIMLIPMLLIFLNTGVSALISEKLVSADETWVQTAKIIGSTPIALLISVLVALFVLGRK 300
 50 orfl40-1.pep RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMA DLGIPVLLGC 360
 orfl40a RGECSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMA DLGIPVLLGC 360
 55 orfl40-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDG 420
 orfl40a FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDG 420
 60 orfl40-1.pep FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI V 461
 orfl40a FWLVGRLLDM DVPTTLKTWT VNQTLIALIG FALSALLFAI V 461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

10	1	MDGRTOTLSA	OTLLGISAAA	IILILILIVK	FRIRALITLV	IASLLTALAT
	51	GLPTGSIVND	VLVKNFGGTL	GGVALLVGLG	AMLGRLEVTS	GGAQSLADAL
	101	IRMFGEKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPIVFAT	ARRMKQDVLP
	151	FALASVGAFS	VMHVFLPPHP	GPIAASEFYG	ANIGQVILIG	LPTAFITWYF
15	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQSDP	PKEPAKAGTV	VAVMLIPMLL
	251	IFLNTGVSAL	ISEKLVSAD	TWVQTAKMIG	STPVALLISV	LAALLVLGRK
	301	RGESGSTLEK	TVDGALAPAC	SVILITGAGG	MFGGVLRASG	IGKALADSM
	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDQQLA
	401	CIVLATAAGS	VGCSEHNDSC	FWLVGRLSDM	DVPTTLKTWT	VNQTLLIAFIG
	451	FALSALLFAI	V*			

	1	ATGGACGGCC	GGACACAGAC	GCTGTCCGCG	CAAACCTTGT	TGGGCATTTC
	51	GGCGGCGGCA	ATCATCTCTA	TTCTGATTTT	AATCGTCAAA	TCCCGCATCC
	101	GCGCGCTGCT	GACACTGGTC	ATCGCCAGCC	TGCTGACGGC	TTTGGCAACC
25	151	GGTTTGGCCA	CAGGCAGCAT	CGTCAACGAC	GTACTGGTCA	AAAACCTTCG
	201	CGGCACGCTC	GGCGGCGTGG	CGCTTCTGGT	CGGTCTGGGC	GCAATGCTCG
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGCAC	AGTCTGCTGC	GGACGCGCTG
	301	ATCCGGATGT	TCGGCGAAAA	ACGCGCACCG	TCGCTCCGGG	CGCTTGCTCT
	351	GCTGATTTTC	GGCTTCCCGA	TTTTCTTCGA	TGCCGGACTA	ATCGTTCATG
	401	TGCCCATCGT	ATTCGCCACC	GCACGGCGCA	TGAAACAGGA	CGTACTGCCC
30	451	TTCCGCGCTT	GCTCCGTCGG	CGCATTTTCC	GTCATGCACG	TCTTCTTGCC
	501	GCCCATCCG	GGCCCGATTG	CCGCTTCCGA	ATTTTACGGC	GGCAACATCG
	551	GCCAGGTTTT	GATTTTGGGT	CTGCCGACCG	CCTTCATCAC	ATGGTATTTC
	601	AGCGGCTATA	TGCTCGGCAA	AGTGTGGGG	CGCGCCATCC	ATGTTTCCCGT
35	651	TCCCGAACTG	CTCAGCGGCG	GCACGCAAGA	CAGCGACCCG	CCGAAAGAAC
	701	CTGCCAAAGC	AGGAACGGTC	GTCGCCGTCA	TGCTGATTCC	CATGCTGCTG
	751	ATTTTCTCTA	ATACCGGCGT	ATCAGCCCTC	ATCAGCGAAA	AACTCGTAAG
	801	TGCGGACGAA	ACTTGGGTTC	AGACGGCAAA	AATGATCGGT	TCGACACCTG
	851	TCGCCCTTCT	GATTTTCGTA	TTGGCCGCAC	TGTTGGTCTT	GGGACGCAAA
40	901	CGCGGCGAAA	GCGGCAGCAC	GTTGGA AAAA	ACCGTGAGAC	GCGCACTCGC
	951	CCCCGCTGT	TCCGTGATTC	TGATTACCGG	CGCGGGCGGT	ATGTTTCGGC
	1001	CGGTTTTCG	TCCTTCGGC	ATCGGCAAGG	CAGCTGCGGA	CAGCATGGCG
	1051	GATTTGGGCA	TTCCCGTCTT	TTTGGGCTGC	TTCTTGTGCG	CCTTGGCACT
	1101	GCGTATCGCG	CAAGGTTCCG	CAACCGTCGC	CCTGACCACA	GCCGCCGCGC
45	1151	TGATGGCTCC	TGCCGTTGCC	GCCGCCGGCT	TTACCGACTG	CGAGCTCGCC
	1201	TGATATCGTAT	TGGCAACGGC	GGCAGGTTTC	GTCGGTTGCA	GCCACTTCAA
	1251	CGACTCCGGC	TTCTGGCTGG	TCGGCCGCCT	CTTGGATATG	GACGTACCGA
	1301	CCACGCTGAA	AACTGGGACG	GTCACCAAAA	CCCTCATCGC	ATTTCATCGG
	1351	TTTGCTTGT	CCGCACTGCT	GTTTGCATC	GCTGTA	

50	1	MDGRTQTL	SA QTL	LGISAAA	IILILILIVK	FRIRALLTLV	IASLLTALAT
	51	GLPTGSIVND	VLVKNFEGTL	GGVALLVGLG	AMLGRIVETS	GGAQSLADAL	
	101	IRMFGKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPTIVFAT	ARRMKQDVL	
	151	FALASVGAFS	VMHVFLPPH	GPIAASEFYG	ANIGQVLILG	LPTAFITWYF	
55	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQSDP	PKEPAKAGTV	VAVMLIEMLL	
	251	IFLNTGVSAL	ISEKLVSAD	TWVQTAKMIG	STPVALISV	LAALLVLGRK	
	301	RGESGSTLEK	TVDGALAPAC	SVILITGAGG	MFGGVLRASG	IGKALADSMA	
	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDWQLA	
	401	CIVLATAAGS	VGCSEHND	SG FVLVGRLLDM	DVPTTLTKTWT	VNQTLIAFIG	
	451	FALSALLFAI	V*				

orf140ng-1.pep MDGRTQTLSAQTLGGISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND

-337-

```

orf140-1      MDGWTQTLSAQTLGLGISAAIILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
orf140ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
5 orf140-1      ILVKNFGGTLGGVALLVGLGAMLGRLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
orf140ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASVGAFSVMHVFLPPHPGPIAASEFYG
10 orf140-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPPFALASIGAFSVMHVFLPPHPGPIAASEFYG
orf140ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQSDPPKEPAKAGTV
orf140-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV
15 orf140ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWQTAKMIGSTPVALLISVLAALLVLGRK
orf140-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWQTAKIIGSTPIALLISVLVALFVLGRK
orf140ng-1.pep RGESESTLEKTVGDALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
20 orf140-1      RGESESALEKTVGDALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
orf140ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFENDSG
25 orf140-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFENDSG
orf140ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
orf140-1      FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV

```

Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454
Score = 210 bits (529), Expect = 1e-53
Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

```

```

Query: 88  ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
          E SGGA+SLA+  R  G+KR  A  +A+  G P+FFD G I++ PI++ A+ K
Sbjct: 80  EHSAGAESLANFYSRKLGDKRTIAALTAAFFLGIPVFFDVGFILAPIIYGFYAKVAKIS 139

```

```

Query: 148  VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
          L F L  G  +HV +PPHPGP+AA+  A+IG + I+G+ + I  GY  K
Sbjct: 140  PLKFGPLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

```

```

Query: 208  VLGRAIHVPVPELL-----SGGTQSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
          ++ + +  E+L  G T+ SD  P A V ++++IP+ +I  T
Sbjct: 199  IINKRQYAMSVEVLEQMQLAPASEEGATKLSDKINPPGVA-LVTSILIVIPAIIMAGT-- 255

```

```

Query: 258  SALISEKLVSADETWQTAKMIGSTPXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
          +S L+  + T ++IGS  +RG S  + AL
Sbjct: 256  ---VSATLMPPSHPLIGTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALEP 312

```

```

Query: 318  PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
          A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
Sbjct: 313  TAAVVILVTGAGGVFGKVLVESGVGKALANMLQMIDLPLLPAAFIISLALRASQGS--AT 370

```

```

Query: 378  XXXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFENDSGFWLVGRLLDMDVPTTLK 437
          G  Q  + LA  G +G SH NDSGFV+V + L + V  LK
Sbjct: 371  VAILTTGGLLSEAVMGLNPIQCVLVTLAACFGGLGASHINDSGFWIVTKYLGLSVADGLK 430

```

```

Query: 438  TWTVNQTLIAFIGFALSALLFAIV 461
          TWTV T++ F GF ++ ++A++
Sbjct: 431  TWTVLTTLTGFTGFLITWCVWAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

(double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTGCG CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
     101  GCGTATTTTT TGCCGTTATC GGACTGACTT CCTGCGGCTT TGCCGGTTTC
     151  AACTTTTGG GCAGACACCA CGGGCGCAC. GTCGTCCTGA TTCTCATCGG
     201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
     251  CCTTTGCCCG CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
     301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
     351  GTTGGCAGCA GCTTATCCGG CAGCATTTGC CCTGATGCTG CCCTTGCCCG
     401  TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKHLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAAGL VLHGYSLARL
     101  RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMMFFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTTG TGGCCCGGCG
     101  TGTTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
     151  GTCGAAGCAC TGGCAGGCAG CCCCACCCCC TTGGTTGCCC ATCTGTTCGG
     201  TCAAACCGAT TTCGGCATA CCGCCGTGTA TCTTTGGGTT GCCGCCGCGT
     251  TCAAACATTT GCTGTGCGCG TGGGCTGCCG ACTCATACGA TGCCGCACGC
     301  TTTGCAGGCG TATTTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC
     351  CCGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
     401  TCATCGGCTG TATCGGGCTG ATTCCAGTTG CCCATTCTCT CAACCCCGCT
     451  GCCGCCGCGT TTGCCGCCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
     501  TCGCCGCGCG GTGATTGCCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
     551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTTGCCCT GATGCTGCCC
     601  TTGCCCGTAC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTGATGTT
     651  GACGGCAGTC GCCTCACTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
     701  CGTGCTCTT GGCAAAAACG CAGCCCGCGC TGTTCGCGCA ATGGCTCGAC
     751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CCGCAGTTC AGACGGCATT
     801  CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGGTTTGCA TTGCCCGCGC
     851  TGCCGCTGGC GGTTTGGACG GTTGCCGCA CGGCCTGTT TTCGACCGAC
     901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
     951  CGTCAATCCG CAGCGTTTTT AGGATAACCT CGTCTGGCTG CTCCGCGCG
    1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
    1051  GCGTTTGTC ACTGGTTCGG CATTATGGCG TTCGGACTGT TTGCCGTGTT
    1101  CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
    1151  CCGAACGCGC GCCTATTTC AGCCCGTATT ATGTTCTCTGA TATCGATCCC
    1201  ATTCCGATGG CGGTGCGCGT ACTGTTTACA CCCTTGTTGGC TGTGGGCGAT
    1251  TACCCGGAAA AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCAG
    1301  GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
    1351  GACGCGGCGA AAAGCCACGC GCCGGTCGTC CGGAGTATGG AGGCATCGCT
    1401  TTCCCCGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
    1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTGGACGCA GTACGGCACA
    1501  TTGCCGCACC GCGTCGGCGA TGTACAATGC CGCTACCGCA TCGTCCTCCT
    1551  GCCCAAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
    1601  CCGGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
    1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLIMAFawl WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLLSP WAADSYDAAR
     101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA
     151  AAAFAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLLAKT QPALFAQWLD

```

Computer analysis of this amino acid sequence gave the following results:

10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
35	51	AAAGCCGTGG	CTGTTGCTGT	TGATGCGGTT	TGCCTGGTTG	TGGCCCGGCG
	101	TGTTTTCCCA	CGATTGTGTG	AATCCTGACG	AACCTGCCGT	CTATACCGCG
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCT	TGTGGTTGCC	ATCTGTTTCGG
	201	TCAAATCGAT	TTCGGCATAC	CGCCCCGTGA	TCTTTGGGTT	GCCGCCGCGT
40	251	TCAAACATTT	GCTGTGCGCG	TGGGCTGCCG	ACCCGTATGA	TGCCGCAGCA
	301	TTTGCCGGCG	TGTTTTTCGC	CGTTGTCCGA	CTGACTTCTT	GCGGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTC	GTCCTGATTG
	401	TCATCGGCTG	TATCGGGCTG	ATTCCGACCG	TACACTTTCT	CAACCCCGCT
45	451	GCCGCCGCCT	TTGCCGCCCG	CGGACTGGTG	CTGCACGGTT	ATTCTTTGGC
	501	TCGCGCGGCG	GTGATTGCCG	CCTCTTTTCT	GCTCGGTACG	GGTTGGACGC
	551	TGATGTCGTT	GGCAGCAGCT	TATCCGGCGG	CATTTGCCCT	GATGCTGCCC
	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
50	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTCTT	GGCAAAAAAC	CAGCCCGCGC	TGTTCCGCGA	ATGGCTCGAC
	751	GATCACGGTT	TCGGTAGGTT	CGCGCGCGTG	CGGCACATTG	ACGACGGCATT
	801	CAGTTTGTGTT	TACTATCTGA	AAAACCTGCT	TTGGTTTGCA	TTGCCTGCGC
55	851	TGCCCGTGGC	GGTTTGACAG	GTTTGCCGCA	CGCGCCTGTT	TTGCACGCAT
	901	TGGGGGATTT	TGGGCGTCGT	CTGGATGCTT	CCCGTTTTTG	TGCTGCTTGC
	951	CGTCAATCCG	CAGCGTTTTT	AGGATAACCT	CGTCTGGCTG	CTTCCGCGCG
	1001	TTGCCCTGTT	CGCGCGGGCG	CAACTGGACA	GCCTGAGACG	CGGCGCGGGC
60	1051	GCGTTTGTCA	ACTGGTTCGG	CATTATGGCG	TTCGGACTGT	TTGCCGTGTT
	1101	CCTGTGGACG	GGCTTTTTTC	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCCTATTTC	AGCCCCGATT	ATGTTCTCTG	ATCGATFCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTTGG	TGTGGGCGAT
60	1251	TACCCGCAAA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTACCTT	GACCTGGGCT	TTGCTGATGA	CGCTGTTCTT	GCCGTGGCTG
	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGCT
	1401	TTCCCCGGAA	TTAAACCGGG	AGCTTTTCAGA	CGGCATCGAG	TGTATCGACA
60	1451	TAGGCGGCGC	CGACCTACAC	ACGCGGATTG	TTTGGACGCA	GTACGGCACA
	1501	TTGCCCGCAC	GCGTCGGCGA	TGTACAATGC	CGCTACCGCA	TCGTCGCGCT
	1551	GCCCCAAAAAC	CGGGATGCGC	CGCAAGGCTG	CGACAGGGTC	TGCGAGGGTG
	1601	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG	CGGCGGCGCG

-340-

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
 1651 GAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

```

5      1 MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR
      101 FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
      151 AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
      201 LPVLMFFRPW QSRRMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD
      251 DHVFGTFGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCTRLESTD
10      301 WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QLDLRRGAA
      351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
      401 IPMAVAVLFT PLWLWAITRK NIRGRQAVTN WAAGVTLTWA LLMTLELPWL
      451 DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGGDLH TRIVWTQYGT
      501 LPHRVGDVQC RYRIVRLPON ADAPQGWQTV WQGARPRNKD SKFALIRKTG
15      551 ENILKTTD*
  
```

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

```

      orf141a.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
      orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20      orf141a.pep LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
      orf141-1 LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN
25      orf141a.pep FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT
      orf141-1 FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT
30      orf141a.pep GWTLMSLAAAYPAAAFALMLPLPVLMFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
      orf141-1 GWTLMSLAAAYPAAAFALMLPLPVLMFFRPWQSRRMLTAVASLAFALPLMTVYPLLAKT
35      orf141a.pep QPALFAQWLDDHVFGTFGGVRHIQTAFSLFYLLKNLLWFALPALPLAVWTVCTRLESTD
      orf141-1 QPALFAQWLDYHVFGTFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCTRLESTD
40      orf141a.pep WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
      orf141-1 WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
45      orf141a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAITRK
      orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAITRK
50      orf141a.pep NIRGRQAVTNWAAGVTLTWALLMTLELPWLDAAKSHAPVVRSMELKRELSDGIE
      orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLELPWLDAAKSHAPVVRSMELKRELSDGIE
55      orf141a.pep CIDIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPONADAPQGWQTVWQGARPRNKD
      orf141-1 CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPONADAPQGWQTVWQGARPRNKD
      orf141a.pep SKFALIRKTGENI
      orf141-1 SKFALIRKIGENI
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

```

60      orf141.pep DFGISPVYLWVAAAFKHLLSPWAADSYDVA 30
      orf141ng WNPAEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADHPYDAA 126
  
```

	orf141.pep	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRXVVLILIGCIGLIPVAHFLNPAAAAFAAAGL	90
	orf141.ng	RFAGVFFAVIGLTSCGFAGFNFLGRHHGRSVVLIHIGCIGLIPVAHFFNPAAAAFAAAGL	186
5	orf141.pep	VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRP	140
	orf141.ng	VLHGYSLARRRVIAASFLLGTGWTLMSLAAAYPAAFALMLPLPVLMMFFRPWQSRRLMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARP	LCEYLLHLAI	RPFLTLMLT	YTPPDARPPA	KTHEKPWLLL
	51	<u>LMFAFWLWPG</u>	<u>VFSHDLWNPA</u>	<u>EPAVYTAVEA</u>	<u>LAGSPTPLVA</u>	<u>HLFGQTDGFI</u>
	101	<u>PPVYLWVAAA</u>	<u>FKHLLSPWAA</u>	<u>HPYDAARFAG</u>	<u>VFFAVIGLTS</u>	<u>CGFAGFNFLG</u>
	151	<u>RHHGRSVVLI</u>	<u>HIGCIGLIPV</u>	<u>AHFFNPAAAA</u>	<u>FAAAGLVLHG</u>	<u>YSLARRRVIA</u>
	201	<u>ASFLLGTGWT</u>	<u>LMSLAAAYPA</u>	<u>AFALMLPLPV</u>	<u>LMFFRPWQSR</u>	<u>RLMLTAVASL</u>
15	251	<u>AFALPLMTVY</u>	<u>PLLLAKTQPA</u>	<u>LFAQWLNHYV</u>	<u>FGTFGGVRHI</u>	<u>QRAFSLPHYL</u>
	301	<u>KNLWFWAPP</u>	<u>LPLAVWTVCR</u>	<u>TRLFSTDWGI</u>	<u>LGIVWMLAVL</u>	<u>VLLAFNPQRF</u>
	351	<u>QDNLVWLLPP</u>	<u>LALFGAAQLD</u>	<u>SLRRGAAAFV</u>	<u>NWFGIMAFGL</u>	<u>FAVFLWTGFF</u>
	401	<u>AMNYGWPAPL</u>	<u>AERAAFYSPY</u>	<u>YVPDIDPIPM</u>	<u>AVAVLFTPLW</u>	<u>LWAITRKNIR</u>
	451	<u>GRQAVTNWAA</u>	<u>GVTLTWALLM</u>	<u>TLFLPWLDAA</u>	<u>KSHAPVVRSM</u>	<u>EASFSPELKR</u>
20	501	<u>ELSDGIEICIG</u>	<u>IGGGDLHTRI</u>	<u>VWTQYGTLPH</u>	<u>RVGDVRCRYR</u>	<u>IVRLPQNADA</u>
	551	<u>PQGWQTVWQG</u>	<u>ARPRNKDSKF</u>	<u>ALIRKIGENI</u>	<u>LKTTD*</u>	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTGCTGT	TGATGGCGTT	TGCCTGGCTG	TGGCCCGGCG
25	101	TGTTTCCCA	CGATTTGTGG	AATCCTGCCG	AACTGCGCGT	CTATACCGCC
	151	GTCGAAGCAC	TGGCAGGCAG	CCCCACCCCG	TTGGTTGCCG	ATCTGTTCCG
	201	TCAAACCGAT	TTCGCGCATAC	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTCGCCG	TGGGCAGCCG	ACCCGTATGA	TGCCGCACGC
	301	TTTTCAGGCG	TATTTTGTGC	CGTTATCGGA	CTGACTTCTT	GCGGCTTTGC
30	351	CGGTTTCAAC	TTTTTGGGCA	GACACCACGG	GCGCAGCGTT	GTTTAAATCC
	401	ATATCGGCTG	TATCGGGCTG	ATTCGGGTG	CCCATTTTCT	CAATCCcgcc
	451	gccgcgcgct	tTGC CGCCG	CGGACTGGTG	CTGCacggt	actcgctgGC
	501	ACGCCGGCGC	GTGATtgccg	cctctTtccT	GCTCGGTACG	GGTTGGACGT
	551	TGATGTTCGT	GGCGCGACGT	TATCCGGCGG	CGTTTGCCTG	GATGCTGCCG
35	601	CTGCCCGTGC	TGATGTTTTT	CCGTCCGTGG	CAAAGCAGGC	GTTTGATGTT
	651	GACGGCAGTC	GCCTCGCTTG	CCTTTGCCCT	GCCGCTTATG	ACCGTTTACC
	701	CGCTGCTCtt	gGCAAAAACG	CAGCCCGCGC	TGTTTGC GCA	ATGGCTCAAC
	751	TATCACGTTT	TCCGTACGtt	cggcgGCGTG	CGGCACA tTC	AGAggGCatT
	801	Cagtttgttt	cactatctgA	AAaatctgct	ttggttcgca	ccgcccgggC
40	851	TGCCGCTGGC	GGTTTGGACG	GTTTGCCGCA	CACGCTGT	TTCGACCGAC
	901	TGGGGGATTT	TGGGCATTGT	CTGGATGCTT	GCCGTTTTGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTC	AAGACAACCT	CGTCTGGCTG	CTGCCGCCGC
	1001	TTGCCCTGTT	CGGCGCGGCG	CAACTGGACA	GCCTGAGGCG	CGGCGCGGCG
	1051	GCTTTTGTCA	ACTGGTTCCG	CATTATGGCG	TTCCGGCTGT	TTGCCGTGTT
45	1101	CTGTGTGACG	GGCTTTTTCG	CCATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CCGAACGCGC	CGCTACTTTC	AGCCCGTATT	ACGTTCCCGA	CATCGATCCC
	1201	ATTCCGATGG	CGGTTGCCGT	ACTGTTTACA	CCCTTGTGGC	TGTGGGCGAT
	1251	TACCCGGA	AACATACGCG	GCAGGCAGGC	GGTTACCAAC	TGGGCGGCAG
	1301	GCGTTACCC	GACCTGGGCT	TTGCTGATGA	CGCTGTTTCT	GCCGTGGCTG
50	1351	GACGCGGCGA	AAAGCCACGC	GCCCGTCGTC	CGGAGTATGG	AGGCATCGTT
	1401	TTCCCGGAA	TTAAACGGG	AGCTTTCAGA	CGGCATCGAG	TGTATCGGCA
	1451	TAGGCGGCGG	CGACCTGCAC	ACGCGGATTG	TTTGACGCA	GTACGGCACA
	1501	TGTCGCGACC	GCGTCGGCGA	TGTCGTTTGC	CGCTACCGTA	TCGTCCGCT
	1551	GCCCCAAAC	GCGGATGCGC	CGCAAGGCTG	GCAGACGCTC	TGGCAGGCTG
55	1601	CGCGCCCGCG	CAACAAAGAC	AGTAAGTTTG	CACGTATACG	GAAAATCGGG
	1651	GAAAATATAT	TAAAACAAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFAWL	WPGVFSHDLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPPVYLWV	AAAFKHLSP	WAADPYDAAR
60	101	<u>FAGVFFAVIG</u>	<u>LTSCGFAGFN</u>	<u>FLGRHHGRSV</u>	<u>VLIHIGCIGL</u>	<u>IPVAHFLNPA</u>
	151	<u>AAAFAAAGLV</u>	<u>LHGYSLARRR</u>	<u>VIAASFLLGT</u>	<u>GWTLMSLAAA</u>	<u>YPAAAFALMLP</u>
	201	<u>LPVLMFFRPW</u>	<u>QSRRLMLTAV</u>	<u>ASLAFALPLM</u>	<u>TVYPLLLAKT</u>	<u>QPALFAQWLN</u>
	251	<u>YHVFGTFGGV</u>	<u>RHIQRAFSLF</u>	<u>HYLKNLLWFA</u>	<u>PPGLPLAVWT</u>	<u>VCRTRLFSTD</u>
	301	<u>WGILGIVWML</u>	<u>AVLVLLAFNP</u>	<u>QRFDQNLVWL</u>	<u>LPPLALFGAA</u>	<u>QLDSLRRGAA</u>

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
 401 IPMAVAVLET PLWLWAIIRK NIRGRQAVTN WAAGVTLTWA LLMTLFLFWL
 451 DAAKSHAPVV RSMEASFSPK LKRELSGIE CIGIGGGDLH TRIVWTQYGT
 501 LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
 551 ENILKTTD*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

orf141ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAPAVYTAVEALAGSPTP
 orf141-1 MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
 orf141ng-1.pep LVAHLFGQTDGFI PPVYLWVAAAFKHLSPWAADPYDAARFAGVFFAVIGLTSCGFAGFN
 orf141-1 LVAHLFGQTDGFI PPVYLWVAAAFKHLSPWAADSYDAARFAGVFFAVIGLTSCGFAGFN
 orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAFAAFAAAGLVLHGYSIARRRVIAASFLLGT
 orf141-1 FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAFAAFAAAGLVLHGYSIARRRVIAASFLLGT
 orf141ng-1.pep GWTLMSLAAAYPAAFALMLPLPVMFFRPWQSRRLMTAVASLAFALPLMTVYPLLLAKT
 orf141-1 GWTLMSLAAAYPAAFALMLPLPVMFFRPWQSRRLMTAVASLAFALPLMTVYPLLLAKT
 orf141ng-1.pep QPALFAQWLNHYHVFGEFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCRTRLFSTD
 orf141-1 QPALFAQWLDYHVFGEFGGVRHVQTAFSLFYLYLKNLLWFALPALPLAVWTVCRTRLFSTD
 orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
 orf141-1 WGILGVVWMLAVLVLLAVNPQRFDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
 orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAIIRK
 orf141-1 FGLFAVFLWTGFFAMNYGWPAPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAIIRK
 orf141ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLFWLDAKSHAPVVRSMESFSPKRELSGIE
 orf141-1 NIRGRQAVTNWAAGVTLTWALLMTLFLFWLDAKSHAPVVRSMESLSPELRELSGIE
 orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLPHRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
 orf141-1 CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD
 orf141ng-1.pep SKFALIRKIGENILKTTDX
 orf141-1 SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

1 ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
 51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CTGCATTAC GATATATTGA
 101 CCGGCCCGGC ATTGAAAAAG CCCGAATTTT TCCAATCAAG GAAATGGGCA
 151 AGCGGTTTTC AGGTAGGCTA TACGTTTTAA

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

1 ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
 51 SGFQVGYTF*

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GGCACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAAATT
 101 ATGGACGTTT GATTGGCGGT ACGCCCGATG AGGAAAGTTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 5 201 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT
 301 ACTGATTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCGGT GTAAAACGTG GGATGAGGGA AACAAAAAGT TACATTGATG
 10 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CTGCGGGTTG GTTGGCAGAA
 451 CTTTCCCACA AAGAATATAT CGGTGCGAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACGC GGCACCGGCA TGAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTCAAAT CGGTAAACAG CTATTTCCTT ATGACACATC
 15 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGTCGGCAC AGCAATTGGG
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
 20 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA
 1001 GCGGTTTCA GGTAGGCTAT ACGTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
 25 51 RKEGSSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNKSYN
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
 151 LSHKEYIGRS TADFKLYKGR GTGMKDALRA PEEAFGEGTS RMKIWTASAD
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPPLTSQ DKLAIGGHHT VRGFDGEMSL
 251 SAERGWWYRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLVGTAIG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35 orf142.pep QSAKWLSGQTLVGTAIGIRGQIKLGGNLHY 30
 orf142ng RGWYWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59
 40 orf142ng DIFTGRALKKPEYFQTKKWVTFQVGYSYF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GGCACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTT TATGTAAATT
 45 101 ATGGACGTTT AATTGGCGGT ACGCCCGATG AGGAAAATTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 201 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC
 301 ACTGATTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCAGT GTAAAACGTG GGACGAGGGA AACAAAAAGT TACATTGATG
 50 401 ATGCCGAAC TACTGTACAA CGGCGTAAAA CCACAGGTTG GTTGGCAGAA
 451 CTTTCCCACA AAGGATATAT CGGTGCGAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACAC GGCACCGGCA TGAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTCAAAT CGGTAAACAG CTATTTCCTT ATGACACATC
 55 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAAACTGG
 701 CTATCGGCGG ACACCACACC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 CTTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGCCGGCAC AGCAATTGGG
 60 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

-344-

1001 CGGGGTTTCA GGTGGGTTAT TCGTTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQA VSGLSE VYDYNKSYN
 101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE
 151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGTS RMKIWTASAD
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
 251 PAERGWWYRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KVVTFQVGY SF*

10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

15 orf142-1.pep MDNSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
 orf142ng-1 MDNSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA
 20 orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLG
 orf142ng-1 VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLS
 25 orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAELSHKEYIGRSTADFKLKYKRGTMKDALRA
 orf142ng-1 VKLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALRA
 30 orf142-1.pep PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
 orf142ng-1 PEEAFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
 35 orf142-1.pep VRGFDGEMSLSAERGWWYRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAG
 orf142ng-1 VRGFDGEMSLPAERGWWYRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG
 orf142-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
 orf142ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKVVTFQVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E.chrysanthemi*:

40 gi11772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
 Score = 119 bits (295), Expect = 3e-26
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
 Query: 2 DNSGSEATGKYQGNITFSA DNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
 DMSG ++TG+ Q N + + DN FGL+D +++ G S + + D + G
 Sbjct: 230 DMSGQKSTGEEQLNGSLALDNVFGADQWFISAGHS---SRFATSHDAESLQAG----- 280
 45 Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGFNRLLYRDAKRKTYLSV 121
 +S P+G W +N++ RY + G S F +R+++RD KT ++
 Sbjct: 281 -FSMPYGYWNLGYNYSQSRYNRTFINRDFPWHSTGDS DTHRFSLSRVVFRDGTMTKTAIAG 339
 50 Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAELSHKGYIGRSTADFKLKYKHGTGMKDALARAP 181
 R +Y++ + L RK + ++H + A F Y G +
 Sbjct: 340 TFSQRTGNNYLNGSLLPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNVRGVRWLGSSETDT 399
 55 Query: 182 EEAFFGEGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++
 Sbjct: 400 DKSADPRAEFNKWTLASYYHPV---TDSITYLGSGLYQYSARALYGSEQLTLGGESSI 456
 Query: 242 RGFDGEMSLPAERGWWYRNDLSWQFKP---GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296
 RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G
 60 Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515
 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKVVTFQVGYYSF 342
 A+G+ + L + G + P + Q V G++VG SF

5 Example 73

10

1	ATGCGGACGA	AATGGTCAGC	AGTGAGAAGC	TGCTTACTTG	GgCGGACACC
51	GCCGACATCG	ATACCGCTTT	GAACCTGTTG	TACCGTTTGC	AAAACTCGA
101	ATTCCTCTAT	GCGCATGAAA	ACGGTCATTG	AGACGGTCAT	AATTTGwCGG
151	ACGAGCAATT	CGCGTTGCTG	ATGGAACAAT	TGTCGGGCAG	CGGTAAAGCG
201	TTATTGGTCG	ATCGGAACGG	TCTGTATCTT	GCCAACGCCA	ATTTCCATCA
251	TGAGGCGGCG	GAAAGAGTTG	GGTTGTTGGC	GCAGAAGTC	GCACAGATGG
301	AAAAGAAATA	CCGGCTGCTG	ATTAAGAACA	GC..	

```

15      1  MRTKWSAVRS  CTWADTADID  TALNLLYRLQ  KLEFLYGDEN  GHSDGINLXD
      51  EQLPLLMEQL  SGSGKALLVD  RNGLYLANAN  FHHEAAEELG  LLAAEVAQME
     101  KKYRLLIKNN  ..

```

20	1	ATGGAATCAA	CACTTTCACT	ACAAGCAAAT	TTATATCCCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGCCCCCACT	GCCGGTAAAA
	101	CTTTGTGCA	CAGCCGTGTG	AAAGCAGATG	CGGACGAAAT	GGTCAGACAGT
	151	GAGAAGCTGC	TTACTTGGGC	GGACACCGCC	GACATCGATA	CCGCTTTGAA
25	201	CCGTGTTGTAC	CGTTTGCAA	AACTCGAATT	CCCTATGGC	GATGAAAAACG
	251	GTCATTGAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG
	301	GAACAATTGT	CCGGCAGCGG	TAAGGCGTTA	TTGGTCGATC	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGCGGAA	GAGTTGGGGT
30	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAA	AGAAATACCG	GCTGCTGATT
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGGCGTTT	GCGATCCTTC
	501	CGGTGAGAGC	GAATTGACAT	TTTTCCCAT	GTATATCGGT	TCAACCAAAAT
	551	TTATTTTGGT	TGATCGGCGC	ATTCGCCATT	TGGGCAAGAA	GGCATTTGTT
	601	ACTTTGGTAA	GGATTTTATA	CCGCCGTTAC	AGCAACCGCG	TGTAA

35

1	MESTLSLQAN	LYPRLTPAGA	FYAVSSDAPS	AGKTLHLSLL	KADADEMVSS
51	EKLLTWADTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRNGLYLA	NANFHEAAE	ELGLLAAEVA	QMEKKYRLLI
151	KNNLYINNNA	WGVCDPGSGS	ELTFFPLYIG	STKFILVIGG	IPDLGKEAFV
201	TLVRILYRRY	SNRV*			

Homology with a predicted ORF from *N.meningitidis* (strain A)

[illegible]

-346-

```

orf143a      YGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
              80      90      100      110      120      130

5  orf143.pep      100      110
      VAQMEKKYRLLIKNN
      ||||| |||
orf143a      VAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFPLYIGSTKFILVIGGIPDLGKEA
              140      150      160      170      180      190

```

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

```

10      1  ATGGAATCAA CANTTTCACT ACAAGCAAAT TTATATCNCC GCCTGACTCC
      51  TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGNCCCCAGT GCCGGTAAAA
     101  CTTTGTTCGA CAGCCTGTTG AAAGCGGATG CGGACGAAAT GGTNAGCAGT
     151  GAGAAGCTGC TTACCTGGGC GGANACCGCC GACATCGATA CCGCTTTGAA
     201  CCTGTTGTAC CGTTTGCAAA AACTCGAATT CCTCTATGGC GATGAAAACG
15      251  GTCATTCAGA CGGCATCAAT TTGTGGGACG AGCAATTGCC GTTGCTGATG
     301  GAACAATTGT CCGGCAGCGG TAAGGCGTTA TTGGTCGATC GGAACGGTCT
     351  GTATCTTGCC AACGCCAATT TCCATCATGA GCGGCGGAA GAGTTGGGT
     401  TGTGGCGGCG AGAAGTCGCA CAGATGGAAA AGAAATACCG GCTGCNNATT
     451  AAGAACAACC TGTATATCAA CAATAACGCT TGGGCGTTC GCGATCCTTC
20      501  CGGTCAGAGC GAATTGACAT TTTTCCCAT GTATATCGGT TCAACCAAAT
     551  TTATTTTGGT TATCGGCGGC ATTCCCGATT TGGGCAAAGA GGCATTTGTT
     601  ACTTTGGTAA GGATNTTATA CCNCCNGTTA CAGCAACCGC GTGTAAAACT
     651  TGGGAGAGAG GANGGGTTAT GCAGCAATTA TTGA

```

This encodes a protein having amino acid sequence <SEQ ID 614>:

```

25      1  MESTXSLOAN LYXRLTPAGA FYAVSSDXPS AGKTLHSLK KADADEMVSS
     51  EKLLTWAXTA DIDTALNLLY RLQKLEFLYG DENGHS DGIN LSDEQLPLLM
    101  EQLSGSGKAL LVDRNGLYLA NANFHHEAAE ELGLLAAEVA QMEKKYRLXI
    151  KNNLYINNNA WGVCDPSGQS ELTFFPLYIG STKFILVIGG IPDLGKEAFV
    201  TLVRXLYXXL QQPRVKLGRE XGLCSNY*

```

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

```

      orf143a.pep      MESTXSLOANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSEKLLTWAXTA
      orf143-1         MESTLSLOANLYPRLTPAGAFYAVSSDAPSAGKTLHSLKADADEMVSSEKLLTWADTA
35      orf143a.pep      DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
      orf143-1         DIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
40      orf143a.pep      NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNNLYINNNAWGVCDPSGQSELTFFFPLYIG
      orf143-1         NANFHHEAAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPSGQSELTFFFPLYIG
      orf143a.pep      STKFILVIGGIPDLGKEAFVTLVRXLY
45      orf143-1         STKFILVIGGIPDLGKEAFVTLVRILY

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

```

50      orf143.pep      MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLXDEQLPLLMEQL      60
      orf143ng         MRTKWSAVRSCSRADTADIDTALNLLYRLQKLEFLYG DENGHS DGINLSDEQLPLLMEQL      60
55      orf143.pep      SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLLIKNN      110
      orf143ng         SGSGKALLVDRNGLYLANANFHHEAAEELGLLAAEVAQMEKKYRLIRNNLYINNNAWGV      120

```

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

-347-

1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLSD
 51 EQLPLLMEQL SGSGKALLVD RNgLYLANAN FHESAEELG LLAAEVAQME
 101 KKYRLLIRNN LYINNNAWGV CDPGQSELT FFPLYIGSTK FILVIAGIPD
 151 LSKGGICYFG KDFIPPLQQP RVKLGTTGGIM RQLLISILED LNNTSTDIIA
 5 201 SAVISTDGLP MATMLPSHLN SDRVGAISAT LLALGSRVQ ELACGELEQV
 251 MIKKGSGYIL LSQAGKDAVL VLVAKETGRL GLILLDAKRA ARHIAEAI*

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

1 ATGGAATCAA CACTTTCAC TACAAGCGAAT TTATATCCCT GCCTGACTCC
 51 TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
 10 101 CTTTGTTCG CAGCCTGTTG AAAGCGGATG CGGACGAAGT GGTACAGCAGT
 151 GAGAAGCTGC TCGCGGCGGA CACCGCCGAC ATCGATACCG CTTTGAACCT
 201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAAACGGTC
 251 ATTCAGACGG CATCAATTG TCGGACGAGC AATTGCCGTT GCTGATGGAA
 301 CAATTGTCCG GCAGCGGTAA GGCATTATTG GTCGATCGGA ACGGTCTGTA
 15 351 TCTTGCCAAC GCCAATTTCC ATCATGAGTC GGCGGAAGAG TTGGGGTTGT
 401 TGGCGGCAGA AGTCGCACAG ATGGAAGA AATACCGGCT GCTGATTAGG
 451 AACCAACTGT ATATCAACAA TAACGCTTGG GCGCTTTCG ATCCTCCGG
 501 TCAGAGCGAA TTGACATTTT TCCCATTTGA TATCGGTTCA ACCAAATTTA
 551 TTTTGGTTAT CGCCGGCATT CCCGATTGA GCAAAGAGGC ATTTGTTACT
 20 601 TTGGTAAGGA TTTTATACCG CCGTTACAGC AACC CGTGT AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

1 MESTLSLQAN LYPCLTPAGA FYAVSSDAPS AGKTLRLSLL KADADEVVSS
 51 EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQLPLLME
 101 QLSGSGKALL VDRNGLYLAN ANFHESAE LGLLAAEVAQ MEKKYRLLIR
 25 151 NNLYINNNAW GVCDPSGQSE LTFFPLYIGS TKFILVIAGI PDLKSKEAFVT
 201 LVRILYRRYS NRV*

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLA-ADTA 59
 30 orf143-1 MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADADEVVSSEKLLTWADTA 60
 orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 119
 35 orf143-1 DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 120
 orf143ng-1.pep NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 179
 orf143-1 NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 180
 40 orf143ng-1.pep STKFILVIAGIPDLKSKEAFVTLVRILYRRYSNRV 213
 orf143-1 STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV 214

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is
 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 45 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51 GTTTGCATGG TTCGTCTGCC GCCGCTTTGA TGAAGAACGC GTACCGCAGr
 10 101 CGGCGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
 151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCGGTGTTTC ACCGTGGTTC
 201 GGATTGTTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
 251 ACATGGTGTTC CGACTATATC AATGCGTTTC GCGAGCAGGC GAACCGGCTG
 301 ACGGCAATCG GCAGCGTGAT GCTGTCGTT ACCTCGCTGA TGCTGATTTCG
 55 351 GACGATAGAC AATACGTTCA ACCGCATCTG GACCGGTCAA wTyCCAGCGT
 401 CCGTGGATG..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLRLQGL ADNKAFAW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQITVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCGGTGCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCGGTGTTCC ACCGCTGGTC
201 GGATTTCGTTT GTCTCCTTCG TCAACCAAAC CATGTGCGCG CAGGCGCGCG
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCTG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
15 451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGACG CTTTGTCTGT GGGGGCTGTA CCGCTTCGTG
601 CCAACCGCTC TCGTTCCCGC GCGGCAGGCG TTTGTGCGGG CTTTGGCAAC
20 651 AGCGTTTTGT CTGAAACCG CGCGTCCCT CTTCACTTGG TATATGGGCA
701 ATTTCGACGG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG CCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAATATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAGAAGG CAAAGCCTTG CCTGTTCAGG AGTTCAGACG
25 951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGCGGATT CGATTGAGTT GAACGAATC TCAAGCTCT TCGTTTACCG
1101 TCCGTTGCCT GTGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30 1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLRLQGL ADNKAFAW FVRRFDEER VPQAAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
35 151 LSLGVGISEF VGSVQDAALA SGAPQWSGAL RTAATLTFT LLLWGLYRFV
201 PNRFPVPAQA FVGALATAFC LETARSLETW YMGNFDGYRS IYGAFAAVPF
251 FLWLNLNLT LVLGGAVLTS SLSYWQGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLARHGYYI SGROGWVLT
351 GADSIENLNL FKLFFVYRPLP VERDHVNQAV DAVMTPLQOT LNMTLAEFDA
40 401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

meningitidis:

```

45      10      20      30      40      50      60
orfl44.pep MTFLLRLQGLADNKAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
orfl44a    MTFLLRLQGLADNKAFAWFVRRFDEERVVPQAAASMTFTLLALVPVLTVMVAVASIF
50      10      20      30      40      50      60
orfl44.pep PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
orfl44a    PVFDRWSDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
55      70      80      90      100     110     120
orfl44.pep NTFNRIWRVXXQRPWM
orfl44a    NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISEFXVGSVQDAALASGAPQWSGAL
60

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1   ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5   51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTTC ACCGNTGGTC
201 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCT
10 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCAGCGTTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAN CTTTATGACG CTTTTGCTGT GGGGGCTGTA CCGCTNCGTG
15 601 CCAAACCGCT TCGTCCCCTG GCGGCANGCG TTTGTGCGGG CTTTGCAAC
651 AGCGTTCTGT CTGGAACCGC CGCGTTCCCT CTTTACTTGG TATATGGGCA
701 ATTTTCGACG CTACCGCTCG ATTTACGGNG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGNCCT
20 851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGCGAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAACTC TTCAAGCTCT TCGTTTACCG
25 1101 TCCGTTGCCG GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1   MTFLQRLQGL ADNKICAFW FVVRREFDEER VPQAAASMTF TTLLALVPVL
30  51  TVMVAVASIF PVFDRWSDSF VSFVNQTIIV PQADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSXMLIRTID NTFNRIWRVN SQRPWMMQFL VYWALLTFGP
151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
201 PNRFPVPAKA FVGALATAFC LETARSLFTW YMGNF DGYS IYGAF AAVPF
251 FLLWLNLLWT LVLGGAVLTS SLSYWQGEAF RRFDSRGRF DDVLKILLLL
35 301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
351 GADSIELNEL FKL FVYRPLP VERDHVNQAV DAVMMPCLQT LNMTLAEFDA
401 QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40  orf144a.pep  MTFLQRLQGLADNKICAFWFVVRREFDEERVQAAASMTFTTLLALVPVLTVMVAVASIF
    orf144-1    MTFLQRLQGLADNKICAFWFVVRREFDEERVQAAASMTFTTLLALVPVLTVMVAVASIF

    orf144a.pep  PVFDRWSDSFVSFVNQTIIVPQADMVFDYINAFREQANRLTAIGSVMLVVT SXMLIRTID
45  orf144-1    PVFDRWSDSFVSFVNQTIIVPQADMVFDYINAFREQANRLTAIGSVMLVVT SLMLIRTID

    orf144a.pep  NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
50  orf144-1    NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL

    orf144a.pep  RTAATLXFMTLLLWGLYRXVPNRFVPARXAFVGALATAFCLETARSLFTWYMGNF DGYS
    orf144-1    RTAATLTFMTLLLWGLYRFVPNRFVPARQAFVGALATAFCLETARSLFTWYMGNF DGYS

55  orf144a.pep  IYGAF AAVPFLLWLNLLWT LVLGGAVLTS SLSYWQGEAFRRRFD SGRGRFDDVLKILLLL
    orf144-1    IYGAF AAVPFLLWLNLLWT LVLGGAVLTS SLSYWQGEAFRRRFD SGRGRFDDVLKILLLL

    orf144a.pep  DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL
60  orf144-1    DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIELNEL

    orf144a.pep  FKL FVYRPLPVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 408
65  orf144-1    FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFWVVRFFDEERVVPQXAASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWVVRFFSEERVVPQAAASMTFTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144ng	PVFDRWSDSFVSFVNQITVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSLMLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMQFLVYWALLTFGLPSLGVGISFMVGSVQDSVLSGGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
20	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQ	FVGALITAF	LETARFLFTW	YMGNFDDYRS	IYGAFAAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
25	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTFCLQT	LNMTLAEFDA
	401	QAKKQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
30	51	ATTTCATG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTCGC	TTCGATTTTC	CCCGTGTTCC	ACCGCTGGTC
	201	GGATTCGTTC	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGCGCGCG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
35	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTCC
	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCTC	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCCT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTATG	GTCGGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
40	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGCTGTA	CCGCTTCGTG
	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTGGGAG	CTTTGATTAC
	651	GGCATTCTCG	CTGGAGACGG	CACGTTTCCT	GTTCACTTGG	TATATGGGCA
	701	ATTTGCGACG	CTACCGCTCG	ATTTACGGCG	CATTTGCCGC	CGTGCCGTTT
	751	TTCTGCTGT	GGTTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
45	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAATCCT	GCTGCTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCTTG	TCCGTTCAGG	AGTTCAGACG
	951	GCATATCAAT	ATGGGTTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCct	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgccgtG	TTTGACAGACT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQWQGL	ADNKICAFW	FVIRRFSEER	VPQAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSFVNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLV	TSLMLIRTID	NAFNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPAQ	FVGALITAF	LETARFLFTW	YMGNFDDYRS	IYGAFAAVPF
	251	FLLWLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLKT

351 GADSIELSEL FKL FVYRPLP VERDHVNQAV DAVMTPCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5      orf144ng-1.pep MTF LQRWQGLADNKICAFWFVIRRFSEERVPOAAASMTFTLLALVPVLTVMVASIF
      orf144-1       MTF LQRWQGLADNKICAFWFVVRRFDEERVPOAAASMTFTLLALVPVLTVMVASIF
      orf144ng-1.pep PVFDRWSDSFVSFVNQTI V PQADMVF DYIDAFRDQANRLTAIGSVMLVVTSMLIRTID
10     orf144-1       PVFDRWSDSFVSFVNQTI V PQADMVF DYINAFREQANRLTAIGSVMLVVTSMLIRTID
      orf144ng-1.pep NAFNRIWRVNTQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
      orf144-1       NTFNRIWRVNSQRPWMMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
15     orf144ng-1.pep KTAARLAFMTLLWGLYRFVFNRFVPAQAFV GALITAFCL ETARFLFTWYMGNF DGYS
      orf144-1       RTAATLTFTLLWGLYRFVFNRFVPAQAFV GALATAFCL ETARSLFTWYMGNF DGYS
20     orf144ng-1.pep IYGAF AAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILL
      orf144-1       IYGAF AAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILL
      orf144ng-1.pep DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGWVLKTGADSIELSE
25     orf144-1       DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYYIYSGRQGWVLKTGADSIELNEL
      orf144ng-1.pep FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS
30     orf144-1       FKL FVYRPLPVERDHVNQAVDAVMT PCLQTLNMTLAEFDAQAKKRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

      1 ..AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
      51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
40    101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
      151 ACCCGCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
      201 CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

      1 ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
      51 TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

      1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
      51 CGAACGCTAC CGTACCGCC GCCTCATCCA CGCGTCCGG CTCGGCGGGG
      101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
      151 GAGTGGATAG GGATGACCGT CTTGTCGTC CTCGGCATGC TCCAGTTTCA
50    201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
      251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACAGCA TTATTTCCAC
      301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
      351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
      401 CGATGTGTAT GCTCATCGGC GACAACGGCA GCGAATGGCT CGACAGCGGA
55    451 CTCATGCGCG CCATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC

```


-352-

501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAAGC CCTCGAGGAG AACATGGCGA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCATCTCGCC GCCACATCGG
 5 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATCG TCAACACCAC CGAGTGCTC CTGACCACCG CCGCCAAGCT
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 10 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGAFLFATA SARLLHLQHG
 51 EWIGMTVFVV LGMLQFOGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCMLIG DNGSEWLD SG
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
 20 251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLLE TREHG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

				10	20	30
	orf146.pep			RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF				
		280 290 300 310 320 330				
		40 50 60 70				
35	orf146.pep	LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHRLRQSLLTREHGX				
		:				
	orf146a	LWLSTNMRQEISALVILLQRTTRRKWLDAHERQHRLRQSLLTREHSX				
		340 350 360 370				

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCGGG CTCGGCGGGG
 101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
 151 GAGTGGATAG GGATGACCGT CTTCTGCTGC CTCGGCATGC TCCAGTTTCA
 201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGG ACGGTCATCG
 45 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACAGCA TTATTCCAC
 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGTT CGACAGCGGC
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
 50 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAAGC CCTCGAAGAG AACATGGCGA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCACCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATCG TCAACACCAC CGAGTGCTC CTGACCACCG CCGCCAAGCT
 55 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 60 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

      1  MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
    51  EWIGMTVFVV LGMLQFQGA IYKKAVERMLG TVIGLGAGLG VLWLNQHYFH
   101  GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWFD SG
    5  151  LMRAMNVLIG AAIATAA AKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
   201  RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
   251  RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
   301  RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
   351  TRRKWLDAHE RQHLRQSLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

      orf146a.pep  MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
      orf146-1     MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
   15  orf146a.pep  LGMLQFQGA IYKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVTASALAGWAA
      orf146-1     LGMLQFQGA IYKKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVTASALAGWAA
   20  orf146a.pep  VGKNGYVPM LAGLTMCLIGDNGSEWFD SGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
      orf146-1     VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAKLLPLKSTLMWR
   25  orf146a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
      orf146-1     FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
   30  orf146a.pep  AMMEAMQHAHRKIVNTTELLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
      orf146-1     AMMEAMQHAHRKIVNTTELLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
   35  orf146a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
      orf146-1     RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
   40  orf146a.pep  RQHLRQSLE TREHSX
      orf146-1     RQHLRQSLE TREHGX

```

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

      orf146.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
      orf146ng   KLNSEIRLLDRHFTLLQTDLQQTAA LINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
   45  orf146.pep  LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLE TREHG 75
      orf146ng   LWLSTNMRQEISALVIPLQTRRKWLDAHERQHLRQSLE TREHG 409

```

50 An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

      1  MSGVRFPSA PIPSTDPPSG SLCFFTFPLQ TASDMNSSQR KRLSGRWLNS
    51  YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGEWIGMT VFVVLGMLQF
   101  QGAIYSNAVE RMLGTVIGL AGLGVWLWNQ HYFHGNLLFY LTIGTASALA
   55  151  GWAAGVKNKY VPM LAGLTMCLIGDNGSEW LD SGLMRAMN VLIGAAIAIA
   201  AAKLLPLKST LMWRFMLADN LADCSKMIAE ISNGRRMTRE RLEQNMVKMR
   251  QINARMVKS RSHLAATSGE RISPSMMEAM QHAHRKIVNT TELLTTAAK
   301  LQSPKLNSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
   351  EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQTRRKWL DAHERQHLRQ
   401  SLE TREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

-354-

1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
 51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCggaa
 101 ccgtCCTGTT CGCCACCGCA CTCGCCCCGgc tACTCCACCT CCAacacggc
 151 gAATGGATAG GGATgaCCGT CTTGTCGTC CTCGGCATGC TCCAGTTCCA
 5 AGGCgcgatt tActccaacg cggtgGAacg taTgctcgt acggtcatcg
 251 ggctgGGCGC GGGTTTGGGc gTTTATGTC TGAACCAGCA TTAttccac
 301 ggcaacCTcc tcttctacct gaccatcg caggcaagcg cactggccgg
 351 ctGGGCGGCG GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
 10 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCCG
 501 CGCAAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGCGGTATGA CGCGCGAAGC TTTGGAGCAG AATATGGTCA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCACCTCGCC GCCACATCGG
 15 CGCAAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 801 GCATCTCCC AACTCAACG GCAGCGAAAT CCGCTGCTC GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
 901 AGACACGCCG GCCGCATCCG CATCGACACC GCCATCAACC CCGAATGGA
 20 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTCCGCCC TCGTCATCCT GCTGCAACGC
 1051 ACCGCGCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

1 MNSSQRKRLS GRWLSYERY RHRRLHAVR LGGTVLFATA LARLLHLQHG
 51 EWIGMTVFVV LGMLQFQGA IYNAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM AGLTMCMLIG DNGSEWLD SG
 151 LMRAMNVLIG AATAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSILA ATSGESRISP SMMEAMQHAH
 251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTAAALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRKWLDAHE RQHLRQSLE TREHG*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSRWLSYERYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
 orf146ng-1 MNSSQRKRLSGRWLSYERYRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
 orf146-1.pep LGMLQFQGA IYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 40 orf146ng-1 LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 45 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSILAATSGESRISP
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNMVKMRQINARMVKSRSILAATSGESRISP
 50 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTVALING
 orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNSEIRLLDRHFTLLQTDLQQTAAALING
 55 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQRTTRRKWLDAHE
 orf146-1.pep RQHLRQSLE TREHGX
 60 orf146ng-1 RQHLRQSLE TREHGX

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

sp|P33011|YEEA_ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
 65 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
 ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi11788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

Query: 20 YRHRRLIHAVRLGGTVLGFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIYSNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHYRIVHGTRVALAFLTLFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCLMI 139
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
Sbjct: 75 GTVLGSILGLIALQLE---LISLPIMLVWCAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

Query: 140 GDNGSEWLDGLMRAMNVLIGXXXXXXXXKLPLKSTLMWREMLADNLADCSKMIAEISN 199
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRSBGDVLGSLAMLFTGIWPQRAFIHWRIQLAKSLTEYNRVYQSAFS 190

Query: 200 GRRMTRERLEQNVMKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V
Sbjct: 191 PNLLEPRLESHLQKLL---TDAVKMRGLIAPASKETRIKPSIYEGIQITINRNLVCMLEL 247

Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAALINGRHARRIRIDTAINPEL 316
+ LN ++R D AL G +N +
Sbjct: 248 QINAYWATRPSPHFVLLNAQKLR--DTQHMMQQILLSLVHALYEGNPQPVFANTEKLNDAV 305

Query: 317 EALAEHL--HYQWQ-----GFLWLSTNMRQEISALVILLQRTTRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLNNHHDLVKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the
gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

```

1  ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
51  GGGCAAATC GTCAGTGTGC GCGAACACAA CGAACGCGAG ATGGCGGACA
101 AGATTGTCCG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCGGAT
151 GCGGGTACGC CGGCCGTGTG CGACCCGGGC GCGAAACTCG CCCGCCGCGT
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATT TTTATTTCAAC
301 GGTTTTGTAC CGCCGAAATC GGGAGAACGC AGGAAACTGT TTGCCAAATG
351 GGTGCGGGCG GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG
401 GTGCAGCGCT TGCCGATATG GCGGAACTGT TCCCCGAACG CCGATTATG
451 CTGGCGCGCG AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT
501 TGGGGAAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
45  551 AGATGGTGTT GGTGCTTTAT CCGGCGCAGG ATGAAAACA CGAAGGCTTG
601 TCCGAGTCCG CGCAAACAT CATGAAATC CTCACAGCCG AGCTGCCGAC
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GAAAGAAAG
701 CTTTGTACGA T..

```

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

```

1  ..AEDTRVTAQL LSAYGIQGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51  AGTPAVCDPG AKLARRVREA GFKVVPVVA XAVMAALSVA GVEGSDFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAALADM AELFPERRIM
151 LAREITKTFE TFLSGTVGEI QTALSADGQ SRGEMVLVLY PAQDEKHEGL
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKKALYD..

```

Further work revealed the complete nucleotide sequence <SEQ ID 641>:

```

1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101 CGCGTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAG GCAAACCTCGT

```

-356-

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85

```

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGGCT
251 ATCTTTTCTAGA CGGCATGGTT GTGGCACAGG TTTCGGATGC GGGTACGCCG
301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401 GCGTGGCCGG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTGTACCG
451 CCGAAATCGG GAGAACGCAG GAAACTGTT GCCAAATGGG TCGGGCGCGC
501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551 CCGATATGGC GGAACGTGTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
601 ATTACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGGTGT GGGAAATTCA
651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGCGGAG ATGGTGTGTTG
701 TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751 CAAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCCGC
801 GGAGCTTGCT GCCAAATCA CGGCGGAGG AAAGAAAGCT TTGTACGATC
851 TGGCTCTGTC TTGGAAGAAC AAATAG

```

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

20
25
30
35
40
45
50
55
60
65
70
75
80
85

```

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
51 RVTAQLLSAY GIQGLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLARE
201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251 QNIMKILTAE LPTKQAELA AKITGEGKKA LYDLALSWKN K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25
30
35
40
45
50
55
60
65
70
75
80
85

```

Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
Orf286: 43 AEDTRHTGLLLQHFGINARLFAHLDHNEQQKAETLLAKLQEGQNTALVSDAGTPLINDPG 102

Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKAVVRA 120
L R RE F + GF+P KS RR
Orf286: 103 YHLVRTCREAGIRVPLPGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAE 162

Orf147: 121 AFPIVMFETPHRIGAAALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
Orf286: 163 PRTLIFYESTHRLDLSLEDIVAVLGESRYVVLARELTKTWETIHGAPVGEALLAWVKEDEN 222

Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAELA AKITGEGKKALY 236
+ +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAEPLPKAAALAAEIHGVKKNALY 278

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

```

orf147.pep AEDTRVTAQLLSAYGIQGLVSVREHNERQ
|||
orf75a TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ
20 30 40 50 60 70

orf147.pep MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA
|||
orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA
80 90 100 110 120 130

orf147.pep GVEGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGAAALADMAELFPERRLM
|||
orf75a GVAGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGATLADMAELFPERRLM
140 150 160 170 180 190

orf147.pep LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
160 170 180 190 200 210

```

```

|||||
orf75a      LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
           200      210      220      230      240      250

           220      230
orf147.pep  LTAELPTKQAAELAAKITGEGKKALYD
|||||
orf75a      LTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX
           260      270      280      290

```

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
45	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGCCAGAC	ATTACCTTGC
	101	CGCCTTTGGC	GGTATTGCAA	AAGCGGACAA	TGCATTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTTCAGG	CGAGGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAA	GTAATCGGTT
	251	TCCTTTTCAGA	CGGCCCTGGT	GTGGCGCAGG	TTTCCGATGC	GGGTACGCCG
50	301	CGCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
	451	CCGAATCTGC	GCGAACGTAG	GAAATTGTTT	CGCAAATGGG	TGCGGCGCGC
	501	ATTTCTGTG	GTCTATGTTG	AAACGCCGCA	CCCAATCGGG	GCAACGCTTG
55	551	CCGATATGGC	GGAATTGTTT	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCACGAAAA	CGTTTGAAC	GTTCTTAAAG	GGCAGCGTTG	GGGAATTACA
	651	GACGCGATTG	GCGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGGG
	701	TGCTTTATCC	GCGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCGA	TGAAAATCCT	TGCGGCGGAG	CTGCCGACCA	AGCAGGCGGC
60	801	GGAGCTTGCC	GCCAAAGATTA	CAGGTGAGGG	CAAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	TGGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGLTYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
    51 RVTQQLLSAY GIQGRVSVR EHNERQMAVK VIGFLSDGLV VAQVSDAGTP
   101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
    5 151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLDMAELF PERRMLLARE
   201 ITKTFFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
   251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*

```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

10  sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
   15  Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

    Query: 4   KHLQKASDSVVGGLTYVVATPIGNLADITLRALAVLQKADIICAEDTRVTQQLLSAYGIQ 63
           K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
    Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLLQHFGIN 59

   20  Query: 64  GRVSVREHNERQMAVKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
           RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
    Sbjct: 60  ARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

   25  Query: 124 VGASAVMAALSAGVAESDFYFNGFVPPKSGERRKLFKQWVRAAFPVVMFETPHRIGATL 183
           G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
    Sbjct: 120 PGPCAAITALSAGLPSDRFCYEGFLPAKSKGRRDALKATIEAEPRTLIFYESTHRLDLSL 179

   30  Query: 184 ADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
           D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
    Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238

   35  Query: 243 HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
           E L A + +L AELP K+AA LAA+I G K ALY AL
    Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
    51 AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
   101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC ACACTTATTT CGGCATCAAC
   45 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
   201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
   251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
   301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
   351 CGCGCTATAA CAACGTTGAT TTTGGTGCGG AAGGAak.AA tATCCC.GAT
   401 CAACAWCGww TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
   50 451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGCATA
   501 AATwTGTCAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
   551 CGGAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
   601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCGCGAAA
   651 GTTCATATCA TATTGCAAGT .....
   701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
   751 AAAGTGGTTA ATTAATGGGG TATTGCAAAAC GGGCAACCCC TATATAGGAA
   801 AAAGCAATGG CTTCCAGCTG GTTCGTAAAG ATTGGTTCTA TGATGAAATC
   851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACGTC AAAATGGGAA
   901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC

```

-359-

5 951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
 1051 AGGTGGTGTG AACAGTTATC GACCCAGACT GAATAATGGA GAAAAATTTT
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA
 1201 AAATAACGAA ACTTGGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
 1251 CCGTTACTTG GAAAGTAAAC GCGTGGCAA ACGACCGCCT GTCCAAAATC
 1301 GGCAAAGGCA CGCTG.....
 10 2101
 2151 TGACTGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
 2201 GATCACGCTC ATTTAAATCT CACAGGGCTT GCCACACTCA ACGGCAATCT
 2251 TAGTGCAAAT GGCATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
 2301 ACGGCAACCK TAAGCCTCGtG G.sAATGcCC AAGCAACATT TAATCAAGCC
 15 2351 ACATTAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCGGC AACGCTAAGG
 2451 CAAACGTAAG CCATTCCGCA CTCACCGTA ATGTCTCCCT AGCCGATAAG
 2501 GCAGTATTCC ATTTTGAAG CAGCCGCTT ACCGGACAAA TCAGCGGCGG
 2551 CAAGGATACG GCATTACACT TAAAGACAG CGAATGGACG CTGCCGTCAg
 20 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT
 2651 TCCGCTATC GCCACGATGC GGCAGGGCG CAAACCGGCA GTGCGACAGA
 2701 TGGCGCGCGC CGCGTTCGC GCCGTTTCGG CCGTTCCTTA TTATmCGTTA
 2751 CACGCCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAAACGGC
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA
 25 2851 CCCGAGCGAC AAATTGAAGC TGGCGGAAAG TTCGAAGGC ACTTACACCT
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG
 2951 GTAGTGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC
 3001 CTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....
 30 3551
 3601 CCGCAACGCC GTTGGACAA GCGGCATCCG GGACACAAA CACTACCGTT
 3651 CGCAAGATTT CCGCGCCTAC CGCCAACAAA CCGACCTGCG CCAATCGGT
 3701 ATGCAGAAAA ACCTCGGCAG CCGGCGCGTC GGCATCCTGT TTTGCGACAA
 35 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG
 3801 CCACGCGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC
 3851 ATCAGCGCGC GCGCGGGT TTAGCAGCG CAGCCTTTcA GACGGCATCG
 3901 GAGsmAAAwT CCGCGCCGC GTGctGCATT ACGGCATTCA GGCACGAtAC
 3951 CGCGCGgtt tCggCGgAtt CCGCATCGAA CCGCACATCG GCGCAACGCG
 40 4001 ctATTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
 4051 CCCCCGCGCT TGCAATCAAC CGcTACCGC CGGGCATTAa GGCAGATTAT
 4101 TCATTCAAAC CCGCGCAACA CATTTCCATC ACGCTTATT TGAGCCTGTC
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCG
 4201 TATTGGCTCA GGATTTTCGG AAAACCGCA GTGCGGAATG GGgCGTAAAC
 45 4251 CCGGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG
 4301 CCGCAACTG GAAGCGCAAC ACAGCGCGG CATCAAATTA GGCTACCGCT
 4351 GTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

50 1 MKTTDKRTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN
 51 YQYRDFAEEN KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG
 101 VAALVGQYI VSAHNGGYN NVDFGAEGXN IXDQXRTYK IVKRNNYKAG
 151 TKGHYPGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDVRVIGA
 201 GRQYWRSEDEPNNRESSYH IAS.....GS PMFIYDAQKQ
 251 KWLINGVLQT GNPIYKSNQ FQLVRKDWFY DEIFAGDTHS VFYEPRQNGK
 301 YSFNDNNGT GKINAKHEHN SLPNRLKTRT VQLENVSLSE TAREPVYHAA
 55 351 GGVNSYRPRL NNGENISFID EGKGLILTS NINQGAGGLY FQGDFTVSPE
 401 NNETWQAGAV HISEDSTVTW KVNQVANDRL SKIGKGTLL.....
 701DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
 60 801 DHAVQNGSLT LSGNAKANVS HSLNNGVSL ADKAVFHFE SRTGQISGG
 851 KDTALHLKDS EWTLPSPGXL GNLDNATI TLNSAYRHDA AGAQTGSATD
 901 APRRRSRSR RSLXVTPPT SVESRFNTLT VNGKLNGQGT FRFMSELFY
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT
 1001 LQNEHVDAGA W.....
 65 1151
 1201 RNAVWTSGR DTKHYRSQDF RAYRQQTDLR QIGMOKNLGS GRVGILFSHN
 1251 RTENTFDDGI GNSARLAHGA VFGQYIDRF YIGISAGAGF SSGSLSDGIG
 1301 XKGRRRVLHY GIQARYRAGF GGFIEPHIG ATRYFVKAD YRYENVNIAT
 70 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDASG KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

```

5      1  ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51  AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTAGCCATA TGCCTGTCGT
      101  TCGGCATTCT TCCCCAAGCC TGGCGGGGAC AACTTATTT CGGCATCAAC
      151  TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
      201  GGCGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAT
      251  CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
10     301  GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
      351  CGGCTATAAC AACGTTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
      401  ATCGTTTTAC TTATAAAATT GTGAAACGGA ATAATTATAA AGCAGGGACT
      451  AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCCGGTT TGCATAAATT
      501  TGTACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
15     551  AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCTGAT TGGGGCAGGC
      601  AGGCAATATT GGCATCTGA TGAAGATGAG CCAATAAACC GCGAAAGTTC
      651  ATATCATATT GCAAGTGGCT ATTCTTGGCT CGTTGGTGCG AATACCTTTG
      701  CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTAGGTAG TGAAAAAATT
      751  AACATAGCC CATATGGTTT TTACCAACA GGAGGCTCAT TTGGCGACAG
20     801  TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
      851  ATGGGGTATT GCAAACGGGC AACCCTATA TAGGAAAAAG CAATGGCTTC
      901  CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
      951  CCATTCACTA TTCTACGAAC CAGTCAAAA TGGGAAATAC TCTTTAACC
25    1001  ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
      1051  CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
      1101  ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
      1151  GTTATCCACC CAGACTGAAT AATGGAGAAA ATATTTCTTT TATTGACGAA
      1201  GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
      1251  ATTATATTTC CAAGGAGATT TTACGGTCTC GCCTGAAAAT AACGAACTTT
30     1301  GGCAAGCGCG GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGAAA
      1351  GTAAACGGCG TGGCAAACGA CCGCTGTCC AAAATCGGCA AAGGCACGCT
      1401  GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
      1451  GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
35     1501  TTTAGTGAAA TCGGCTTGGT CAGCGGCAGG GGTACGGTGC AACTGAATGC
      1551  CGATAATCAG TTCAACCCCG ACAAACTCTA TTTGGGCTTT CGCGGCGGAC
      1601  GTTTGGATTT AAACGGGCAT TCGCTTTCGT TCCACCGTAT TCAAAATACC
      1651  GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
      1701  TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACAACAGCT
40     1751  TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
      1801  ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACCAGC CCGCCGCGAG
      1851  AGACCGCACC CTGCTGCTTT CCGCGGAAC AAATTTAAAC GGCAACATCA
      1901  CGCAAAACAA CGGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
      1951  TACAATCATT TAAACGCCA TTGGTCGCAA AAAGAGGGCA TTCCTCGCGG
45     2001  GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATT AAAGCGGAAA
      2051  ACTTCCAAAT TAAAGGCGGA CAGGCGGTGG TTTCCCGCAA TGTTGCCAAA
      2101  GTGAAAGGCG ATTGGCATTG GAGCAATCAC GCCAAGCAG TTTTGGTGT
      2151  CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
      2201  TGACAAATTG TGTGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
50     2251  TTGACTAAGA CCGACATCAG CGGCAATGTC GATCTTGCCG ATCACGCTCA
      2301  TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAAATG
      2351  GCGATACACG TTATACAGTC AGCCACAACG CCACCAAAA CCGCAACCTT
      2401  AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
      2451  CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
55     2501  TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGGC AAACGTAAGC
      2551  CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
      2601  TTTTGAAAGC AGCCGCTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
      2651  CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTGAGG CACGGAATTA
      2701  GGCAATTTAA ACCTTGACAA CGCCACCATT ACACTCAATT CCGCTATCG
60     2751  CCACGATGCG GCAGGGGCGC AAACCGGCAG TGCGACAGAT GCGCCGCGCC
      2801  GCGGTCGCG CCGTTCGCGC CGTTCCTTAT TATCCGTTAC ACCGCCAACT
      2851  TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
      2901  TCAGGGAACA TTCCGCTTTA TGTGGAACCT CTTGCGCTAC CGCAGCGACA
      2951  AATTGAAGCT GGCGGAAAGT TCCGAAGGCA CTTACACCTT GGCGGTCAAC
65     3001  AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGGAAGG
      3051  AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTACC CTGCAAAACG
      3101  AACACGTCGA TGCCGGCGCG TGGCGTTACC AACTCATCCG CAAAGACGGC
      3151  GAGTTCGCCG TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
      3201  CCGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAGGCC
70     3251  TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
      3301  GTTGCCGAAC CGGCCCGGCA GGCAGGCGGG GAAAATGTCG GCATTATGCA

```

-361-

3351 GCGGAGGAA GAGAAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
 3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCACCGC CTTCCCCCGC
 3451 GCCCGCCGCG CCCGCCGGGA TTTGCCGCAA CTGCAACCCC AACCGCAGCC
 5 3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGT'TTGAGTG
 3551 AATTTTCCGC CACGCTCAAC AGCGTTTTTCG CCGTACAGGA CGAATTAGAC
 3601 CGCGTATTTC CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
 3651 GGACACCAA CACTACCGTT CGCAAGATTT CCGCGCCTAC CGCCAACAAA
 3701 CCGACCTGCG CCAATCGGT ATGCAGAAA ACCTCGGCAG CGGGCGCGTC
 10 3751 GGCATCCTGT TTTCGCACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
 3801 CGGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA
 3851 TCGACAGGTT CTACATCGCG ATCAGCGCGG GCGGGGTTT TAGCAGCGGC
 3901 AGCCTTTTAC ACGGCATCGG AGGCAAAATC CGCCGCGCGG TGCTGCATTA
 3951 CGGCATTTCG GCACGATACC GCGCGGTTT CCGCGGATTC GGCATCGAAC
 4001 CGCACATCGG CGCAACGCGC TATTTCGTCC AAAAAGCGGA TTACCGCTAC
 15 4051 GAAACGTCA ATATCGCCAC CCCCGCCTT GCATTCAACC GCTACCGCGC
 4101 GGGCATTAA GAGATTATT CATTCAAACC GCGCAACAC ATTTCATCA
 4151 CGCCTTATT GAGCCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
 4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTCGGCA AAACCCGCGC
 4251 TGCGGAATG GCGTAACAG CCGAAATCAA AGGTTTCAGC CTGTCCCTCC
 20 4301 ACGCTGCCGC CGCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
 4351 ATCAAATTAG GCTACCGCTG GTAA

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA WAGHTYFGIN
 25 51 YQYRDFAE N KGFVAGAKD IEVYNKKGEL VGKSMKAPM IDFSVVS RNG
 101 VAALVG DQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
 151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
 201 RQYWRSD EDE PNNRESSYHI ASAYSWLVGG NTFAQNGSGG GTVNLGSEKI
 251 KHSFYGF LPT GSGFGDSGSP MFIYDAQKQK WLINGVLQTG NPYIGKSNF
 301 QLVKRDW FYD EIFAGDTHSV FYEPRQNGKY SFNDNNGTG KINAKHEHNS
 351 LPNRLKTRTV QLFNVSLSET AREPVYHAAG GVNYSYRPRLN NGENISFIDE
 401 KGKELILTSN INQAGGLYF QGDFTVSPEN NETWQAGVH ISEDSTVTWK
 451 VNGVANDRLS KIGKGT LHVQ AKGENQGSIS VGDGTVILDQ QADDKGKKQA
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSEFHRIQNT
 551 DEGAMIVNHN QKESTVTIT GNKDIAATTGN NNSLDSKKEI AYNWGFGEKD
 35 601 TTKTNGRLNL VYQFAAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
 701 VKGDWHL SNH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
 751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
 801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
 40 851 HSLNGNVSL ADKAVFHES SRFTGQISGG KDTALHLKDS EWTLPSTEL
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLSVTPPT
 951 SVESRFTLT VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
 1001 NTGNEPASLE QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG
 1051 EFRLHNPVKE QELSDKLGA EAKQAEKDN AQSLDALIAA GRDAVEKTES
 45 1101 VAEPARQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
 1151 ARRARRDLPO LQPQPQPQPQ RDLISRYANS GLSEFSATLN SVFAVQDELD
 1201 RVFAEDRRNA VWTSGIRD K HYRSQDFRAY RQQTDLRQIG MQKNLGSGRV
 1251 GILFSHN RTE NTFDDGIGNS ARLAHGAVFC QYGIDRFYIG ISAGAGFSSG
 1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
 50 1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDASGKVR
 1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
 1451 IKLGYRW*

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

10 20 30 40 50 60
 orf1.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPOAWAGHTYFGINYQYRDFAE N
 60 orf1a MKTTDKRTTETHRKAPKTGRIRFSPAYLAIACLSFGILPOAWAGHTYFGINYQYRDFAE N
 10 20 30 40 50 60
 70 80 90 100 110 120
 orf1.pep KGKFAVGA KDIEVYNKKGELVGKSMKAPMIDFSVVS RNGVAALVG VQYIVSAHNGGYN

-362-

orfla
70 80 90 100 110 120

5
orfl.pep
130 140 150 160 170 180
orfla
130 140 150 160 170

10
orfl.pep
190 200 210
orfla
180 190 200 210 220 230

15
orfl.pep
220 230 240 250 260
orfla
240 250 260 270 280 290

20
orfl.pep
270 280 290 300 310 320
orfla
300 310 320 330 340 350

25
orfl.pep
330 340 350 360 370 380
orfla
360 370 380 390 400 410

30
orfl.pep
390 400 410 420 430
orfla
420 430 440 450 460 470

35
orfl.pep
orfla

40
orfl.pep
orfla
480 490 500 510 520 530

45
orfl.pep
orfla

50
orfla
540 550 560 570 580 590

55
orfl.pep
orfla
600 610 620 630 640 650

60
orfl.pep
orfla
660 670 680 690 700 710

65
orfl.pep
440 450 460 470 480
orfla
720 730 740 750 760 770

70
orfl.pep
490 500 510 520 530 540
GDTRYTVSHNATQNGNXLVXNAQATFNQATLNGNTSASGNASFNLSHAVQNGSLTSLG

orf1a	GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTSLSD	780	790	800	810	820	830
orf1.pep	NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGNL	550	560	570	580	590	600
orf1a	NAKANVSHSALNGNVSLADKAVFHFENS RFTGQLSGSKXTALHLKDSEWTLPSGTELGNL	840	850	860	870	880	890
orf1.pep	NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXLVTPTTSVESRFNTLTVNG	610	620	630	640	650	660
orf1a	NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFNTLTVNG	900	910	920	930	940	950
orf1.pep	KLNGQGTFRFMSELFGYRSDKLKLAESSEGTYTTLAVNNTGNEPASLEQLTVVEGKDNKPL	670	680	690	700	710	720
orf1a	KLNXQGTFRFMSELFGYRSDKLKLAESSEGTYTTLAVNNTGNEPVSLDQLTVVEGKDNKPL	960	970	980	990	1000	1010
orf1.pep	SENLNFTLQNEHVDAGAW-----	730	740	750			
orf1a	SENLNFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLGKAEAKKQAEKDNAQS	1020	1030	1040	1050	1060	1070
orf1.pep	-----						
orf1a	LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKRVQADKDSALAKQREAE TRP	1080	1090	1100	1110	1120	1130
orf1.pep	-----						760
orf1a	XTTAFPRARXARRDLPPQPQPQPQPQORDLXSRYANSGLSEFSATLNSVFAVQDELDR	1140	1150	1160	1170	1180	1190
orf1.pep	VFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTEN	770	780	790	800	810	820
orf1a	VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTEN	1200	1210	1220	1230	1240	1250
orf1.pep	TFDDGIGNSARLAHGA VFGQY GIDRFYIGISAGAGFSSGSLSDGIGXKXRRRVLHYG IQA	830	840	850	860	870	880
orf1a	XFDDGIGNSARLAHGA VFGQY GIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVLHYG IQA	1260	1270	1280	1290	1300	1310
orf1.pep	RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI	890	900	910	920	930	940
orf1a	RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX	1320	1330	1340	1350	1360	1370
orf1.pep	SITPYLSLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKP	950	960	970	980	990	1000
orf1a	SITPYXLSYTDAA SGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAAAGKP	1380	1390	1400	1410	1420	1430
orf1.pep	QLEAQHSAGIKLGYRWX	1010	1020				
orf1a	QLEAQHSAGIKLGYRWX	1440	1450				

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

5

10

15

20

25

30

35

40

45

50

55

60

65

70

1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
 51 AACCGGCCGC ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCTGT
 101 TCGGCATTCT TCCCAAGCT TGGGCGGGAC AACTTATTT CGGCATCAAC
 151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
 201 GCGGAAAGAT ATTGAGGTNT ACAACAAAA AGGGGAGTTG GTCGGCAAAT
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG GCGTAACGGC
 301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
 351 CGGTATATAAC AACGTTGATT TTGGTGCGGA AGGAAGNAAT CCCGATCAGC
 401 ACCGTTTTTC TTACCAAAT GTGAAAAGAA ATAATTATAA GCCTGACAAT
 451 TCACACCCTT ACAACGGCGA TTANCATATG CCGCGTTTGC ATAAATTTGT
 501 CACAGATGCA GAACCTGTGC AAATGACGAG TGACATGAGG GGAATACCT
 551 ATTCCGATAA AGAAAAATAT CCCGAGCGTG TCCGCATCGG CTCAGGACAC
 601 CACTATTGGC GTTATGATGA TGACAAACAC GCGGATTTAT CCTACTCCGG
 651 CGCATGGTTA ATTGGCGGCA ATACACATAT GCAGGGTTGG GGAATAATG
 701 TCTACANTTAG TTTGAGCGGC GATGTGCGCC ATGCCAACGA CTATGGCCCT
 751 ATGCCGATTG CAGGTGCGGC AGGCGACAGC GGTTCGCCAA TGTTTATTTA
 801 TGACAAAACA AACAATAAAT GGCTGCTCAA CGGAGTTTTA CAAACCGGCT
 851 ACCCTTATTC CCGCAGGGAA AACGGTTTCC AGCTGATACG CAAAGATTGG
 901 TTTACGATG ACATTTACAG AGGCGATACA CATACCGTCT NTTTTGAACC
 951 GCGCAGTAAC GGACATTTTT CTTTACATC CAACAACAAC GGTACGGGTA
 1001 CGGTAACAGA AACCAACGAA AAGGNTTCCA ATCCAAAGCT TAAAGTACAG
 1051 ACAGTCCGAC TGTTTGACGA ATCTTTGAAT GAACTGATA AAGAACCAGT
 1101 TTACGCGGCA GGGGGTGTTA ATCAGTACCG TCCAAGGTTA AACAACGGTG
 1151 AAAACCTTTC TTTTATCGAT TACGGCAACG GCAAATCAT CTTATCAAC
 1201 AACATCAACC AAGGCGCGGG CGGTTTGAT TTTGAAGGTG ATTTTACGGT
 1251 CTGCGCTGAA AACAACGAAA CGTGCAAGG CGCGGGCGTT CATATCAGTG
 1301 AAGACAGTAC CGTTACTTGG AAAGTAAACG GCGTGGCAAA CGACCGCCTG
 1351 TCCAAATCG GCAAAGGCAC GCTGCACGTT CAAGCCAAAG GGGAAAACCA
 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGGAT CAGCAGGCAG
 1451 ACAGATAAAG CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGGC
 1501 AGGGGTACGG TGCAACTGAA TGCCGATAAT CAGTTCAACC CCGACAACT
 1551 CTATTTCCGC TTTCCGCGCG GACGTTTGGA TTTAAACGGG CATTCGCTTT
 1601 CGTTCACCG TATTCAAAAT ACCGATGAAG GGGCGATGAT TGNCNATCAT
 1651 AATGCCACAA CAACATCCAC CGTTACCAT ACAGGGAATG AAAGTATTAC
 1701 ACAACCGAGT GGTAAAGATA TCAATAGACT TAATTACAGC AAAGAAATTG
 1751 CCTACAACGG TTGGTTTGGC GAGAAAGATA CGACCAAAAC GAACGGGCGG
 1801 CTCACCTTG TTTACCAGCC CGCCGAGAA GACCGCACCC NGCTGCTTTC
 1851 CGGCGGAACA AATTAAACG GCAACATCAC GCAAACAAC GGCAAACTGT
 1901 TTTTCAGCGG CAGACCGACA CCGCACGCCT ACAATCATT AGGAAGCGGG
 1951 TGGTCAAAAA TGGAAGGTAT CCCACAAGGA GAAATCGTGT GGGACAACGA
 2001 CTGGATCNAC CGCACGTTTA AAGCGGAAAA TTTCCATATT CAGGCGGGC
 2051 AGGCGGTGAT TTCCCGCAAT GTTGCCAAAG TGGAAGGCGA TTGNCATTTG
 2101 AGCAATCAGC CCCAAGCAGT TTTTGGTGTG GCACCGCATC AAAGCCATAC
 2151 AATCTGTACA CGTTCGACT GGACNGGTCT GACAAATTGT GTCGAANAAA
 2201 NCATTACCGA CGATAAAGT ATTGCTTCAT TGACTAAGAC NGACNTNAGC
 2251 GGCANTGTNA GNCTNNCCNA TNACGNTNNT TNAANCTCN CNGGGCNTGC
 2301 NNCACNTAAN GGCAATCTTA GTGCAATGG CGATACACGT TATACAGTCA
 2351 GCCACAACGC CACCCAAAAC GGCAACCTTA GCCTCGTGGG CAATGCCCCA
 2401 GCAACATTTA ATCAAGCCAC ATTAAACGGC AACNCATCGG NTTCCGGGCA
 2451 TGCTTCATTT AATCTAAGCA ACAACGCCGC ACAAACGGC AGTCTGACGC
 2501 TTTCCGACAA CGCTAAGGCA AACGTAAGCC ATTCCGCACT CAACGGCAAT
 2551 GTCTCCCTAG CCGATAAGGC AGTATTCCAT TTTGAAAACA GCCGCTTTAC
 2601 CGGACAACTC AGCGGCAGCA AGGANACAGC ATTACACTTA AAAGACAGCG
 2651 AATGGACGCT GCCGTACGGC ACGGAATTAG GCAATTTAAA CCTTGACAAC
 2701 GCCACCATTA CACTCAATTC CGCCTATCGC CACGATGCTG CAGGCGCGCA
 2751 AACCGGCAGN GTGTCAGACA CGCCGCGCGC CCGTTCGCGC CGTTCCCTAT
 2801 TATCCGTTAC ACCGCCAACT TCGGTAGAAT CCCGTTTCAA CACGCTGACG
 2851 GTAACCGGCA AATTGAACNG TCAAGGAACA TTCCGCTTTA TGTGGAAGT
 2901 CTTCCGCTAC CGAAGCGACA AATTGAAGCT GCGGAAAGT TCCGAAGGNA
 2951 CTTACACCTT GGCGGTCAAC AATACCGGCA ACGAACCCGT AAGCCTCGAT
 3001 CAATTGACGG TAGTGGAAGG GAAAGACAAC AAACCGCTGT CCGAAAACCT
 3051 TAATTTACAC CTGCAAAACG AACACGTCGA TGCCGGCGCG TGGCGTTACC
 3101 AACTCATCCG CAAAGACGGC GAGTTCGCC TGCATAATCC GGTCAAAGAA
 3151 CAAAGAGCTT CCGACAACCT CGGCAAGGCA GAAGCCAAA AACAGGCGGA
 3201 AAAAGACAAC GCGCAAAGCC TTGACGCGCT GATTGCGGCC GGGCGCGATG
 3251 CCGCGGAAAA GACAGAAAGC GTTGCCGAAC CGGCCCGGCG GGCAGGCGGG
 3301 GAAATGTCTG GCATTATGCA GGCGGAGGAA GAGAAAAAC GGGTGCAGGC
 3351 GGATAAAGAC AGCGCNTTGG CGAAACAGCG CGAAGCGGAA ACCCGGCCGG
 3401 NTACCACCGC CTTCCCCCGC GCCCGCNGCG CCCGCCGGGA TTTGCCGCAA
 3451 CCGCAGCCCC AACCGCAACC TCAACCCCAA CCGCAGCGCG ACCTGATNAG
 3501 CCGTTATGCC AATAGCGGTT TGAGTGAATT TTCCGCCACG CTCACAGCG
 3551 TTTTCGCCGT ACAGGACGAA TTGGACCGCG TGTTCGCCGA AGACCGCGCG

-365-

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTTCGCA
 3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CTGCGGCCAA ATCGGTATGC
 3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAACCGG
 3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
 5 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCGAC ATCGGCATCA
 3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TNTCAGACGG CATCGGAGGC
 3901 AAAATCCGCC GCCGCGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
 3951 CGGTTTTCGGC GGATTTCGGCA TCGAACCCTA CATCGGCGCA ACGCCTATT
 10 4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
 4051 GGTCTTTCGCT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTTCATT
 4101 CAAACCGCGC CAACACATNT CCATCACNCC TTATTNAGC CTGTCCTATA
 4151 CCGATGCCGC TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
 4201 GCTCAGGATT TCGGCAAAAC CCGCAGTCG GAATGGGGCG TAAACGCCGA
 4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCCGCC AAAGNCCGC
 15 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRG
 101 VAALVGDQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPDN
 151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDKEY PERVRIGSGH
 201 HWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNGVXSLSG DVRHANDYGP
 251 MPIAGAAGDS GSPMFIYDKT NNKWLNLGVL QTGYPYSGRE NGFQLIRKDW
 301 FYDDIYRGDT HTVXFEPNSN GHFSFTSNNN GTGTVTETNE KVSNNPKLVQ
 351 TVRLFDESLN ETDKEPVYAA GGVNQYRPRL NNGENLSFID YNGKLLILSN
 401 NNQAGGLY FEGDFTVSPE NNETWQAGV HISEDSTVTW KVNQVANDRL
 451 SKIGKGLTHV QAKGENQCSI SVGDGTVILD QQADDKGGKQ AFSEIGLXSG
 501 RGTVQLNADN QFNPDKLYFG FRGRLDLNG HSLSFHRIQN TDEGAMIXXH
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
 601 LNLVYQPAAE DRTXLLSGGT NLNGNITQTN GKLFSSGRPT PHAYNHLGSG
 30 651 WSKMEGIPQG EIVWDNDWIX RTFKAENFHI QGGQAVISRN VAKVEGDHXL
 701 SNHAQAVFGV APHQSHICT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
 751 GXVXLXXXXX XXLXGXAXLX GNLSANGDTR YTVSHNATON GNLSLVGNAQ
 801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAKA NVSHSALNGN
 851 VSLADKAVFH FENSRTFGQL SGSKXTALHL KDSEWTLPSG TELGNLNLDN
 35 901 ATITLNSAYR HDAAGAQTGX VSDTPRRRSR RSLSVTPPT SVESRFTLT
 951 VNGKLNQGT FRFMSELFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
 1001 QLTVEGKDN KPLSENLFNT LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE
 1051 QLSDKLGA EAKKQAEKDN AQSLDALIA GRDAEAKTES VAEPARXAGG
 1101 ENVGIMQAE EKRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDLPO
 40 1151 PQPQPQPQPQ PQRDLSRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
 1201 NAVWTSXIRX TKHYRSQDER AYRQQTDLRQ IGMQKNLGS RVGILFSHNR
 1251 TENXFDDGIG NSARLAHGA FGQYIGIRFD IGISTGAGFS SGXLSDGIGG
 1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQKADY RYENVNIAIP
 1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VTRVNTAVL
 45 1401 AQDFGKTRSA EWGVNAEIKG FTLXSHAAAA KGPQLEAQHS AGIKLGYRW*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

		10	20	30	40	50	60
50	orf1a.pep	MKTTDKRTTETHRKAPKTGR	IRFSPAYLAI	CLSFGILPQA	WAGHTYFGIN	YQYYRDFAEK	
	orf1-1						
		10	20	30	40	50	60
55	orf1a.pep	KGKFAVGAKD	IEVYNKKGEL	VGKSMTKAPM	IDFSVVSRRG	VAALVGDQYI	VSAHNGGYN
	orf1-1						
		70	80	90	100	110	120
60	orf1a.pep	NVDFGAEGXNPDQHRFSYQI	VKRNNYKPDNS	-HPYNGDXHMPRLHKFVT	DAEPVEMTSDM		
	orf1-1						
		130	140	150	160	170	179
65	orf1a.pep	NVDFGAEGRNPDQHRFTYKIV	KRNNYKAGTKGHPYGGDYHMPRLHKFVT	DAEPVEMTSYM			
		130	140	150	160	170	180

-366-

5	orfla.pep	180	190	200	210	220	230
	orfl-1	190	200	210	220	230	240
10	orfla.pep	240	250	260	270	280	290
	orfl-1	250	260	270	280	290	
15	orfla.pep	300	310	320	330	340	350
	orfl-1	300	310	320	330	340	350
20	orfla.pep	360	370	380	390	400	410
	orfl-1	360	370	380	390	400	410
25	orfla.pep	420	430	440	450	460	470
	orfl-1	420	430	440	450	460	470
30	orfla.pep	480	490	500	510	520	530
	orfl-1	480	490	500	510	520	530
35	orfla.pep	540	550	560	570	580	590
	orfl-1	540	550	560	570	580	590
40	orfla.pep	600	610	620	630	640	650
	orfl-1	600	610	620	630	640	650
45	orfla.pep	660	670	680	690	700	710
	orfl-1	660	670	680	690	700	710
50	orfla.pep	720	730	740	750	760	770
	orfl-1	720	730	740	750	760	770
55	orfla.pep	780	790	800	810	820	830
	orfl-1	780	790	800	810	820	830
60	orfla.pep	840	850	860	870	880	890
	orfl-1	840	850	860	870	880	890
65	orfla.pep						
	orfl-1						
70	orfla.pep						
	orfl-1						

-367-

		900	910	920	930	940	
	orfla.pep	TELGNLNL	DNATITLNS	AYRHDAA	GAQTGXV	SDTPRRRSRRS---LLSVTPPTS	VESRFRN
	orfl-1	TELGNLNL	DNATITLNS	AYRHDAA	GAQTGSATD	APRRRSRRSRRSLLSVTPPTS	VESRFRN
5		900	910	920	930	940	950
	orfla.pep	950	960	970	980	990	1000
	orfl-1	950	960	970	980	990	1000
10		960	970	980	990	1000	1010
	orfla.pep	1010	1020	1030	1040	1050	1060
	orfl-1	1010	1020	1030	1040	1050	1060
15		1020	1030	1040	1050	1060	1070
	orfla.pep	1070	1080	1090	1100	1110	1120
	orfl-1	1070	1080	1090	1100	1110	1120
20		1080	1090	1100	1110	1120	1130
	orfla.pep	1130	1140	1150	1160	1170	1180
	orfl-1	1130	1140	1150	1160	1170	1180
25		1140	1150	1160	1170	1180	1190
	orfla.pep	1190	1200	1210	1220	1230	1240
	orfl-1	1190	1200	1210	1220	1230	1240
30		1200	1210	1220	1230	1240	1250
	orfla.pep	1250	1260	1270	1280	1290	1300
	orfl-1	1250	1260	1270	1280	1290	1300
35		1260	1270	1280	1290	1300	1310
	orfla.pep	1310	1320	1330	1340	1350	1360
	orfl-1	1310	1320	1330	1340	1350	1360
40		1320	1330	1340	1350	1360	1370
	orfla.pep	1370	1380	1390	1400	1410	1420
	orfl-1	1370	1380	1390	1400	1410	1420
45		1380	1390	1400	1410	1420	1430
	orfla.pep	1430	1440	1450			
	orfl-1	1430	1440	1450			
50		1440	1450				
	orfla.pep	1450					
	orfl-1	1450					
55							
	orfla.pep						
	orfl-1						
60							

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

	orfl	23	FXAAYLAICLSFGILPQAWAGHTYFGINYQYYR	DF	AE	NKGKFAVGAKDIEVYNKKGELVG	82
			F +L C+S GI QAWAGHTYFGI+YQYYR	DF	AE	NKGK F VGAK+IEVYNK+G+LVG	
65	hap	6	FRLNFLTACVSLGIASQAWAGHTYFGIDYQYYR	DF	AE	NKGKFTVGAKNIEVYNKEGQLVG	65
	orfl	83	KSMTKAPMIDFSVSRNGVAALVGQYIVSAHNGGYN	VD	FG	AE	GNIXDQXRXYKIV
			SMTKAPMIDFSVSRNGVAALVG QYIVSAHNGGYN	VD	FG	AE	GN DQ R TY+IV
	hap	66	TSMTKAPMIDFSVSRNGVAALVGQYIVSAHNGGYN	VD	FG	AE	GRN-PDQHRFTYQIV
							124

5
10
15
20

```

orf1 143 KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202
hap 125 KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR 184

orf1 203 QYWRSEDEPNNRESSYHIA----- 222
hap 185 QYWR+D+DE N SSY+++
hap 185 QYWRDKDEETNVHSSYYVSGAYRYLTAGNHTQSGNGNGTVNLSGNVVS PNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTNPNYIGKSNQFQVLRKDWFYDEIFAGDTHSVF 277
hap 245 SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF 304

orf1 278 --YEPRQNGKYSFNDDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334
hap 305 Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A 363

orf1 335 AGGVNSYRPRLNNGENISFIDEGKGELILTSNINQAGAGGLYFQGDFTV-SPENNETWQGA 393
hap 364 A G N Y+PR+ G+NI D+GKG L + +NINQAGAGGLYF+G+F V +NN TWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423
hap 424 GV I +D+TV WKV+ NDRLSKIG GTL 453

```

25 Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30
35
40

```

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSDHAVQNGSLTSL 98
hap 733 DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS 792

orf1 99 GNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGXELGN 158
hap 793 +A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N 852

orf1 159 LNLDNATITLNSAYRHDAAQAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESEFNTLTVN 218
hap 853 L L+N+T+TLNSAY + S+ +AP L T PTS E RFNTLTVN 899

orf1 219 GKLNGQGTFRFMSELFGYRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKP 278
hap 900 GKL+GGGTFF+ F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP 959

orf1 279 LSENLFNFTLQNEHVDAGA 296
hap 960 LS+ L FTL+N+HVDAGA 977

```

45 Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50
55
60
65

```

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTHYRSQDFRAYRQQTDLRQIGMQKNLGSGRVGLFHSNR 60
hap 1135 LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+QK L +GR+G +FSH+R 1194

orf1 61 TENTFDDGIGNSARLAHGAVFQGYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120
hap 1195 ++NTFD+ + N A L + F QY K R+ ++YG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSEKPA 180
hap 1255 + A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P 1314

orf1 181 QHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240
hap 1315 +IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + + 1374

orf1 241 KGPQLEAQSAGIKLGYRW 259
hap 1375 +G QL Q + G+KLGyrw 1393

```

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDGFAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSY	180
	orf1ng	NVDGFAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGGSGSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	289
30	orf1.pep	<u>FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRT</u>	315
	orf1ng	<u>FQLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRT</u>	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDEGKGELILTSNINQGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPRLNNGENISFIDKGKGELILTSNINQGAGGLY	
40	orf1.pep	FQGDFTVSPENNETWQAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKTIITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXLVXNAQATFNQATLNGNTSASGNASFNLSNNA	803
	orf1ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orf1.pep	VQNGSLTSLGNAKANVSHSALGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLDNKANVSHSALGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWT	893
60	orf1.pep	LPSGXEELGNLNDNATITLNSAYRHAAGAQTGSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1ng	LPSGTEELGNLNDNATITLNSAYRHAAGAQTGSAADAPRRRSRRS---LLSVTPPTSVE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFYRSDKLLAESSEGTYYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFYRSGKLLAESSEGTYYTLAVNNTGNEPVSLQLT	1010
70	orf1.pep	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLENLNFNLQNEHVDAGAWRYQLIRKDGELRLHNPVKEQELSDKLGKAGET	1070
75	orf1.pep	// LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGA VFGQYIGIDRFY	1271
	orf1ng	AYRQQTDLRQIGMQKNLGSGRVGILFSHNRTGNTFDDGIGNSARLAHGA VFGQYIGIGRFD	1299

```

orf1.pep      IGISAGAGFSSGSLSDGIGXKRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
               |||||
orf1ng        IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359
5  orf1.pep      RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1391
               |||||
orf1ng        RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1419
10 orf1.pep      AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW 1440
               |||||
orf1ng        AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW 1468

```

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCTTAA
15 51 AACCGGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTCGT
101 TCGGCATTCT GCCCAAGCC CGGGCGGGAC ACATTATT TCGCATCAAC
151 TACCAATACT ATCGCGACTT TGGCGAAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
251 CGATGACGAA AGCCCCGATG ATTGATTTTT CTGTGGTATC GCGTAACGGC
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
20 351 CGGTATATAA AATGTTGATT TTGGTGCGGA GGAAGCAAT CCCGATCAGC
401 ACCGCTTTTC TTACCAAAAT GTGAAAAGAA ATAATTATAA AGCAGGGACT
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAATT
501 TGTACAGAT GCAGAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
25 551 AATACGCTGA TTAAATAAAA TACCCTGATC GTGTTTCAAT CGGAGCAGGC
601 AGACAATATT GCGGCTCTGA TGAAGACGAA CCAATAAACC GCGAAAGTTC
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTAGGTAG CGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
30 801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
851 ATGGGGTATT GCAAACAGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTTCAGT TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAAACG
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAAACATAA ACACTATTCT
1051 CTACCTTATA GATTAAAAAC ACGAACCGTT CAATTGTTTA ATGTTTCTTT
35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACAAA
1201 GGAAGAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT
40 1301 GCGAAGGCGC GGGCGTTTCT ATCAGTGATG GCAGTACCGT TACTTGAAA
1351 GTAAACGGCG TGGCAAACGA CCGCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGGCGACG
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
45 1501 TTTAGTGAAA TCGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCGGCTTT CGCGGCGGAC
1601 GTTTGGATT TGAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
45 1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACTATTACA GGCAATAAAG ATATTACTAC AACCGGCAAT AACAACTACT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
50 1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
1851 GGATCGCACT TTAGTCTTT CCGCGGAAC AAATTAAAC GGCAATATCA
1901 CGCAAAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGCACGCC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCCAACAAG
2001 AGAAATCGTG TGGGACAACG ATTGGATCGA CCGCACATT AAAGCGGAAA
55 2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCGCAA TGTTGCCAAA
2101 GTGGAAGGCG ATTGGCATT AAGCAATCAC GCCCAAGCAG TTTTCGGTGT
2151 CGCACCAGAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCCG ATCAGCTCA
60 2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG
2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCCAAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAACCG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACACGCCG
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AACGTAAGC
65 2551 CATTCGGCAC TCAACGGCAA TGCTCCCTA GCCGATAAGG CAGTATTCCA
2601 TTTTGAAGAA AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTAACCTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCGCGGCC
70 2801 GCGGTCGCG CCGTCCCTTA TTATCCGTTA CGCGCCAAC TTCGCAGAA
2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATGAACG GTCAGGGAAC

```

-371-

5 2901 ATTCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAA GAAAAGACAA
 3051 CACACCGCTG TCCGAAAATC TTAATTTAC CCTGCaaaAc gaacacgtcg
 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCcGc
 3151 CTGCATAATC CGGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGgc
 3201 gggagaaACA GAggccgccT TGACGGCAAA ACAGGCacaa CTTGCCGCCA
 3251 AAcaacaggc ggaAAAAGAC AACgcgcaaa gccttgAcgc gctgattgcg
 3301 gCggggcgca atgccaccga AAAGGCagaa agtgttgccg aaccgGCCCCG
 10 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA
 3401 AACGGGTGCA GCGCGATAAA GACACGCCT TGGCGAAACA GCGCGAAGCG
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCCGCCG
 3501 GGATTGCGCG CAACCCGAGC CCCAACCGCA ACCCCAACCG CAGCGCGACC
 15 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC
 3601 AACAGCGTTT TCGCCGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA
 3651 CCGCCGCAAC GCCGTTTGGA CAAGCGGCAT CCGGGACACC AAACACTACC
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC
 3751 GGTATGCAGA AAAACCTCGG CAGCGGCGCG GTCGCGCATC TGTTTTCGCA
 20 3801 CAACCGGACC GGAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC
 3851 TTGCCACCGG TGCCGTTTTC GGGCAATACG GCATCGGCAG GTTCGACATC
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT
 3951 CAGAGGCAAA ATCCGCGGCC GCGTGCTGCA TTACGGCATT CAGGCAAGAT
 4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG
 25 4051 CGCTATTTTCG TCCAAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
 4101 CACCCCGGGC CTTCGATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
 4151 ATTCATTCAA ACCGGCGCAA CACATTCCA TCACGCCTTA TTTGAGCCTG
 4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC
 4251 CGTATGCGCG CAGGATTTTCG GCAAAACCCG CAGTGGCGAA TGGGGCGTAA
 30 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCAGCTGCG CGCCGCCAAG
 4351 GGGCCGCAAT TGGAAGCGCA GCACAGCGCG GGCATCAAAT TAGGCTACCG
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

35 1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA RAGHTYFGIN
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVSRRG
 101 VAALAGDQYI VSVAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT
 151 NGHPYGGDYH MPRLLHKFVTD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG
 201 RQYWRSDDEE PNNRESSYHI ASAYSWLVG GNTFAQNGSGG GTVNLGSEKI
 251 KHSPYGFLLPT GGSFGDSGSP MFIYDAQKQK WLINGVLOTG NPYIGKSNFG
 301 QLVRRDWFYD EIFAGDTHSV FYEPHQNGKY FFNDNNNGAG KIDAKHKHYS
 40 351 LPYRLKTRTV QLFNVSLSET AREPVYHAAG GVNSYRPRLN NGENISFIDK
 401 KGKELILTSN INQAGGLYF EGNFTVSPKN NETWQAGVH ISDGSVTWTK
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQKKKQA
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNGWFGEKD
 601 ATKTNGLNL NYPPEADRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK
 701 VEGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAETrTIRL RANATQNGNL
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS
 50 851 HSALNGNVSL ADKAVFHFEN SRFTGKISGG KDTALHLKDS EWTLPSTGTEL
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRSL LSVTPPTSSE
 951 SRFNTLTVNG KLNGQGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG
 1001 NEPVSLQLT VVEGKONTPL SENLNFLLQN EHVDAGAWRY QLIRKDGFR
 1051 LHNVPKEQEL SDKLGKAGET EAALTAKQAA LAAKQAAEKD NAQSLDALIA
 55 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA
 1151 ETRPATTAFF RARRARRDLP QPQPQPQPQ QRDLSRYAN SGLSEFSATL
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRD KHYRSQDFRA YRQQTDLRQI
 1251 GMQKNLSSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF QQYIGIRFDI
 60 1301 GISAGAGFSS GSLSDGIRGK IRRRVLYGI QARYRAGFGG FGIEPHIGAT
 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL
 1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAK
 1451 GPQLEAQHSA GIKLGYRW*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN					
5	orf1ng-1						
		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
10	orf1ng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYN					
		70	80	90	100	110	120
	orf1-1.pep	130	140	150	160	170	180
15	orf1ng-1	NVDFGAEGRNPQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAEPVEMTSYM					
		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
20	orf1ng-1	DGRKYIDQNNYPDRVRIGAGRQYWRSDDEPNNRESSYHIASAYSWLVGGNTFAQNGSGG					
		190	200	210	220	230	240
	orf1-1.pep	250	260	270	280	290	300
25	orf1ng-1	GTVNLGSEKIKHSPYGFLPTGGSFGDSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF					
		250	260	270	280	290	300
30	orf1-1.pep	310	320	330	340	350	360
	orf1ng-1	QLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTV					
35							
		310	320	330	340	350	360
	orf1-1.pep	370	380	390	400	410	420
40	orf1ng-1	QLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQGAGGLYF					
		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
45	orf1ng-1	QGDETVSPENNETWQAGVHISEDSTVTKVNGVANDRLSKIGKGTLLHVQAKGENQGSIS					
		430	440	450	460	470	480
	orf1-1.pep	490	500	510	520	530	540
50	orf1ng-1	VGDGTVILDQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPKLYFGFRGGRDLNGH					
		490	500	510	520	530	540
	orf1-1.pep	550	560	570	580	590	600
55	orf1ng-1	SLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDIATTGNNNSLDSKKEIAYNGWFGEKD					
		550	560	570	580	590	600
60	orf1-1.pep	610	620	630	640	650	660
	orf1ng-1	TTKTNGRLNLVYQPAEDRTLILSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQ					
65							
		610	620	630	640	650	660
	orf1-1.pep	670	680	690	700	710	720
70	orf1ng-1	KEGIPRGEIVWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHLNSNHAQAVFGVAPH					
		670	680	690	700	710	720

10

15

20

25

30

35

40

45

50

55

65

70

-374-

1430 1440 1450
 orfl-1.pep KGFTLSLHAAAAGPQLEAQSAGIKLGYRWX
 |||||
 5 orflng-1 KGFTLSLHAAAAGPQLEAQSAGIKLGYRWX
 1440 1450 1460

In addition, ORF1ng shows 55.7% identity with hap protein (P45387) over a 1455aa overlap:

SCORES Init1: 1104 Initn: 4632 Opt: 2680

Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

10 orflng-1.pep MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDFAE
 15 p45387 MKKTVFRLNFLTACISLGIVSQAWAGHTYFGIDYQYYRDFAE
 20 orflng-1.pep KGKFAVGAQDIEVYNKKGELVGSMTKAPMIDFSVVSRRNGVAALAGDQYIVSVAHNGGYN
 25 p45387 KGKFTVGAQNIKYNKQGLVGTSMKAPMIDFSVVSRRNGVAALVENQYIVSVAHNVGYT
 30 orflng-1.pep NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTDAPVEMTSYM
 35 p45387 DVDFGAEGNPNPDQHRFTYKIVKRNNYKGD-NLHPYEDDYHNPRLHKFVTEAAPIDMTSNM
 40 orflng-1.pep DGWKYADLNKYPDRVRIGAGRQYWRSEDEPNNRESSYHIASAYSWLVGGNFQAQNGSGG
 45 p45387 NGSTYSRDKYPERVRIGSGRQFWRNDQDKGD-----QVAGAYHYLTAGNTHNQRGAGN
 50 orflng-1.pep GTVNLGSEKIKHSPYGFPLTGGSGSGSPMFIYDAQKQKWLINGVLQTNPNYIGKSNGF
 55 p45387 GYSYLGSDVRKAGEYGPLPIAGSKGSGSPMFIYDAEKQKWLINGILREGNPFEGKENG
 60 orflng-1.pep QLVKRDWFEYDEIFAGDTHSVFEPHONGKYFFNDNNNGAGKIDAKHKHYSLPYRLKTRTV
 65 p45387 QLVKRSYF-DEIFERDLHTSLYTRAGNGVYTISGNDNGQGSITQKS---GIPSEIK---I
 70 orflng-1.pep QLENVSLSETAREPVYHAA-GGVNSYRPRLNNGENISFIDKKGELILTSNINQAGAGGLY
 75 p45387 TLANMSLPLKEKDKVHNPRYDGPNIYSPRLNNGETLYFMDQKQGSILFASDINQAGAGGLY
 80 orflng-1.pep FEKNFTVSPKNNETWQAGVHISDGSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV
 85 p45387 FEKNFTVSPNSQTWQAGIHVSENSTVTWKVNGVEHDLRLSKIGKGTLLVQAKGENKGS
 90 orflng-1.pep SVGDGKVILDQQADDQGGKQAFSEIGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLN
 95 p45387 SVGDGKVILEQQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDTDKFYFGFRGGRDLN
 100 orflng-1.pep HSLSFHRIQNTDEGAMIVNHNQDKESTVTITGNKDITT-TGNN-NNLDSKKEIAYNGWFG
 105 p45387 HSLSFKRIQNTDEGAMIVNHNQQAANVTITGNESIVLPNGNNINKLDYRKEIAYNGWFG

10

15

20

25

30

35

40

45

50

60

65

70

5

15

20 Example 78

25

30

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

40

45

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 20 30

-377-

```

      orf6.pep                                KVVQFVEKPLRAVVPADSFEPTAQKLNLFK
      orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFK
5      40          50          60          70          80          90
      orf6.pep      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY
      orf6a      AGAATILFYEDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY
10     100        110        120        130        140        150
      orf6.pep      NPLPDAAIKAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
      orf6a      NPLPDAAIKAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
15     160        170        180        190        200

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
20 51  TTCGTTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
101 TCGAACACGC CGTTTTGCAC ACACCTTCTT CGTTCAATTC CCAATCTGCC
151 CGTGTGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTGTG
201 CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTT GAACCGACCG
251 CGCAAAATTT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTAT
301 GAAGATCAAA ATGTCGTCAA AGGTTTGAG GAGCAGTTCC CTGCTTATGC
25 351 CGCCAACTTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
401 CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
451 TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCCGA
501 AAATGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
30 551 CAGGTGAAA GACCTTTGAA CCAGTTGCAG AACGTTTGAA AGTGTTCGGC
601 GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

1  MTRQSLQQA ESRRSIYSLN KNLPVGKDEI VQIVEHAVLH TPSSFNSQSA
51  RVVVLFGEH DKVWQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
101 EDQNVVKGLEQFPAYAAANFPVWADQANAM VQYAVWTTLA AVGVGANLQH
35 151 YNPLPDAIA KAWNIPENWL LRAQMVIGGI EGAAGEKTFE PVAERLKVFG
201 A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

40  orf6a.pep      50          60          70          80          90          100
      orf6-1      TPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
      orf6-1      LRAVVPADSFEPTAQKLNLFKAGAATILFY
      10          20          30
45  orf6a.pep      110        120        130        140        150        160
      orf6-1      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAIA
      orf6-1      EDQNVVKGLEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAIA
50  40          50          60          70          80          90
      orf6a.pep      170        180        190        200
      orf6-1      KAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
      orf6-1      KAWNIPENWLLRAQMVIGGIEGAAGEKTFEPVAERLKVFGAX
55  100        110        120        130

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

	orf6.pep	KVWQFVEXPLRAVVPADSFEPTAQKLNLFK	30
	orf6ng	SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLFK	64
5	orf6.pep	AGAATILFYEDQNVVKGLQEFPAYAAANFPVWADQANAMVQYAVWTTLAAVGVGANLQHY	90
	orf6ng	AGAATILFYEDQNVVKGLQEFPAYAAANFPVWADQANAMVQYAVWTTLAAVGAGANLQHY	124
10	orf6.pep	NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGA	140
	orf6ng	NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEFVAERLKVFGA	174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGT
51	ACGCATGGGA	TTACCCTTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT
101	ATAAGGTGTG	GCAATTTGTC	GAAGACGCGC	TGCGTGCCGT	CGTGCCTGCC
151	GACAGTTTTG	AACCGACCGC	GCAAAAATTG	AAGCTGTTTA	AGGCGGGCGC
201	GGCAACCATT	TTGTTTTATG	AAGATCAAAA	TGTCGTCAA	GGTTTGCAGG
251	AGCAGTCCCC	TGCTTATGCC	GCCAACTTTC	CCGTTTGGGC	GGACCAAGGCG
301	AACGCTATGG	TACAGTATGC	CGTCTGGACG	ACACTTGCCG	CGGTCGGTGC
351	AGGTGCAAAT	CTGCCAACATT	ACAACCCCTT	GCCCGATGTG	GCGATTGCTA
401	AAGCGTGGA	TATTCCTGAA	AACTGGCTGT	TGCGCGCGCA	AATGGTTATC
451	GGTGTATTG	AAGGGGcggc	aggtgaaaaa	gtctttgaac	CCGTTGCgga
501	acgtttgAAA	GTGTTCCGCG	CATAA		

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

1	MAVASNVSLD	MSNPTVLRMG	LPLYIASLRR	GAIYKVWQFV	EDALRAVVA
51	DSFEPTAQKL	KLFKAGAATI	LFYEDQNVVK	GLQEFPAYA	ANFPVWADQA
101	NAMVQYAVWT	TLAAGAGAN	LQHYNPLPDV	AIKAWNIP	NWLLRAQMVI
151	GGIEGAAGEK	VFEPVAERLK	VFGA*		

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

	orf6-1.pep					10	20	30
						LRAVVPADSFEPTAQKLNLFKAGAATILFY		
35	orf6ng	PTVLRMGLPLYIASLRRGAIYKVWQFVEDALRAVVPADSFEPTAQKLKLFKAGAATILFY	20	30	40	50	60	70
	orf6-1.pep		40	50	60	70	80	90
40		EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLPDAAIA						
	orf6ng	EDQNVVKGLQEQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGAGANLQHYNPLPDVAIA	80	90	100	110	120	130
	orf6-1.pep		100	110	120	130		
45		KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX						
	orf6ng	KAWNIPENWLLRAQMVIIGGIEGAAGEKVFEFVAERLKVFGAX	140	150	160	170		

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1	..GGCTACAAC	ACCTGTTCGC	GCGCGGCAGC	CGCATCGCCA	ACTACCAAAT
	51	CAACGGCATC	CCCCTTGCCG	ACGCGCTGGC	CGATACGGG	CAATGCCAAC
	101	ACCGCCGCT	ATGAGCGCGT	AGAAGTCGTG	CGCGGCGTGG	CGGGGCTGCT
	151	GGACGGCAGC	GGCGAGCCTT	CCGCCACCGT	CAATCTGGTG	CGCAAACGCC
	201	TGACCCGCAA	GCCATTGTTT	GAAGTCCGCG	CCGAAGCGGG	CAACCGcAAA

251 CATTTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crC
 301 rCTGCGCGcGC CGCTTGGTTT CCACCTTCGG ACGCGGCGAC TCGTGGCGGC
 351 GCGCGGAACG CAGCGGskAT GCCGAACCTT ACGGCATTTT GGAATACGAC
 401 ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA
 5 451 AGAAACCGCC GACGCGCCGC TCAGcTACGC CGTGTACGAC AGCCAAGGTT
 501 ATGCCACCGC CTTCGGCCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC
 551 AGCCACCACC GTGCGCTCAA CCTGTTGCGC GGCATCGAAC ACCGCTTCAA
 601 CCAAGACTGG AAACCTCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLFFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
 51 DGTGEPSTV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX
 101 LRGRVLSTFG RGDSWRRRER SRXAEYLGIL EYDIAPQTRV HAXMDYQQAQ
 151 ETADAPLSYA VYDSQGYATA FGPKDNPATN WANSHHRLN LFAGIEHREFN
 201 QDWKLKAEYD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATCACC GTTACCGCGC ACCGCACCGC GAGTTCCAAC
 151 GACGGGTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
 20 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAACG CTCGACCGCG CCCTGTTGCA GCGCAGCGGC
 301 ACCAGCCGCC AGATTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
 351 CCGCGCGGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCGTTG
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
 25 451 GTAGAAGTCG TGCGCGCGCT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCTGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGACCGC
 601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGTTCGCGC GCCGCTTGGT
 651 TTCCACCTTC GGACGCGGCG ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
 30 701 ATCCCGAACT CTACGGCATT TTGGAATACG ACATCGCACC GCAAAACCCG
 751 GTCCACGCGC GCATGGACTA CCAGCAGCGC AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
 35 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCATTGAT
 1101 CGGCAAAATAC CGCCTGTTTC GCCGCGAACA CGATTTAATC GCGGGTATCA
 40 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCGC CATCATCCCC
 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 1251 GCCTGCATCG TTTGCCCAAA CCATCCGCGA ATACGGCACC AGGCGCGCAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCCG
 45 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG
 1451 GCATCGTGT CGACCTGACC GGCAACCTGT CTCTTTACGG CTCGTACAGC
 1501 AGCCTGTTCG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGG CATCAAGGC GAATGGCTTG
 1601 AAGGCCGTCT GAACGCATCC GCGCGCTGT ACCGCGCCCG TAAAAACAAC
 1651 TCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 50 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
 1801 GACCAAGACG GCAGCCGCCT GAACCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
 55 1951 ACGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAACTGTCT CTGAACGTGG ACAATCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCGACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
 101 TSRQIYGS DR AGYNLFFARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL LDGTGEPSTV NLVRKRLTR KPLFEVRAEA GNRKHFGLDA
 65 201 DVSGSLNTEG TLRGRVLSTF GRGDSWRRER RSRDAELYGI LEYDIAPQTR

-380-

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQP YGVAGVLSID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYQPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
 5 451 ILGGYTRYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRINAS AAVYRARKNN
 551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHTDPA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVDNLFNKH
 10 701 YRTQPDHSHY GALRTVNAAF TYRFK*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15 Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEP SATVNLVRK 65
 ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
 PupB 215 WSRGFAIQNYEVDGVPTSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273
 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGRLVSTFXXXXXXXXXXXXXAAE 125
 R T + + EAGN +G DVSG L +RGR V+ +
 20 PupB 274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333
 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
 +YGI E+D++ T + Y + D+PL + S G T N A +W+
 25 PupB 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRS LNAAAPDWSY 391
 Orf23 184 SHHRLNLFAGIEHRFNQDWKLKAE 208
 + H + F IE + W K E
 PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N.*

meningitidis:

35 orf23.pep GYNL FARGSRIANYQINGIPVADALADTG
 orf23a QMRDQNIKALDRALLQATGTSRQIYGS DRAGYNL FARGSRIANYQINGIPVADALADTG
 90 100 110 120 130 140
 40 orf23.pep NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD
 orf23a NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLGAD
 150 160 170 180 190 200
 45 orf23.pep VSGSLNTEXXLRGRLVSTFGRGDSWRRRERSR XAELYGILEYDIAPQTRVHAXMDYQQA K
 orf23a VSGSLNAEGTLRGRLVSTFGRGDSWRRRERSR XAELYGILEYDIAPQTRVHAGMDYQQA K
 210 220 230 240 250 260
 50 orf23.pep ETADAPLSYAVYDSQGYATAFGPKDNPATN WANSRHRALNLFAGIEHRFNQDWKLKAEYD
 orf23a ETADAPLSYAVYDSQGYATAFGPKDNPATN WANSRHRALNLFAGIEHRFNQDWKLKAEYD
 270 280 290 300 310 320
 orf23.pep Y
 60 orf23a YTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLF GREHDLIA
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

-381-

1 ATGACACGCT TCAAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCAAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
 201 CCTGCGCGAA ATCCCCGAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GGCGACCGGC
 301 ACCAGCGGCC AGATTTACGG CTCGACCGC GCGGGCTACA ACTACCTGTT
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTTG
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
 10 451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACGGCA AACATTTCGG GCTGGGCGCG
 601 GACGTATCGG GCAGCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCTGTT
 15 651 TTCCACCTTC GGACGCGCG ACTCGTGGCG GCAGCGGAA CGCAGCCGCG
 701 ATGCCGAAC CTACGCGATT TTGGAATACG ACATCGCACC GCAAACCCGC
 751 GTCCACGCG GCATGGACTA CCAGCAGCGC AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 851 CGAAAGACAA CCCCGCCACA AATTGGGCGA ACAGCCGCCA CCGTGCCTC
 901 AACCTGTTTC CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
 20 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGG ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCAATTAAT
 1101 CGGCAAATAC CGCCTGTTTC GCCGGAACA CGATTTAATC GCGGGTATCA
 1151 ACGGTACAA ATACGCCAGC AACAAATACG GCGAACGCG CATCATCCCC
 25 1201 AACGCCATT CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 1251 GCGTGCATCG TTTGCCCAAA CCATCCGCA ATACGCCACC AGGCGGCAAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCGCG CCGCCGACAA CCTTTCGCTG
 1351 ATACTCGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCTACACAG
 30 1451 GCATCGTGT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCTACAGC
 1501 AGCCTGTTTC TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 AACCGTAACC GGCAACAATC TGGAAGCCGG CATCAAAGGC GAATGGCTTG
 1601 AAGCCGCTCT GAACGCATCC GCCGCCGTGT ACCGCGCCCG TAAAAACAAC
 35 1651 CTGCGCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
 1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCC AGCGGCTGGA
 40 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGACC
 1951 ACGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAAGTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GGCGCACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAATAAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLE AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QMRDQNIKA LDRALLQATG
 101 TSRQIYGS DR AGYNYLFARG SRIANYQING IPVADALADT GNANTAYER
 151 VEVVRGVAGL LDGTGEPSAT VNLVRKRPT R KPLFEVRAEA GNRKHFGLGA
 201 DVSGSLNAEG TLRGRIVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKONPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQP YGVAGVLSID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGERSIP
 401 NAIPNAYEFS RTGAYPQPAS FAQTIPQYGT RRQIGGYLAT RFRADNLSL
 55 451 ILGGYRSYR TGSYDSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVPQSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRNLAS AAVYRARKNN
 551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
 601 DQDGSRLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHDPA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
 701 YRTQPDHRSY GALRTVNAAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

orf23a.pep 10 20 30 40 50 60
 MTRFKYSLLEAALLPVYAQADVSVSDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
 |||||
 65 orf23-1 MTRFKYSLLEAALLPVYAQADVSVSDPKPQESTELPTITVTADRTASSNDGYTVSGTHT
 10 20 30 40 50 60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNILFARG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKTLDRALLQATGTSRQIYGSDRAGYNILFARG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGRLVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAPQTRVHAGMDYQQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWNANSRHRAL					
		250	260	270	280	290	300
25	orf23a.pep	310	320	330	340	350	360
	orf23-1	NLFAGIEHRFNQDWKLKAEYDYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPETH					
30		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
35	orf23-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
40	orf23-1	FAQTIPQYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
	orf23a.pep	490	500	510	520	530	540
45	orf23-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLPVPTGNNLEAGIKGEWLEGRINAS					
		490	500	510	520	530	540
	orf23a.pep	550	560	570	580	590	600
50	orf23-1	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
55	orf23a.pep	610	620	630	640	650	660
	orf23-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAK					
60		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
65	orf23-1	ARAADNSRQKAYAVADIMARYRFPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF					
		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N. gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSXAELYGILEYDIAPQTRVHAXMDYQOAKETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQOAKETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWANSRHRALNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRNRALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTKPL	LFEVRAEAGN	RKHFGLGADV
	101	SGSLNAEGTL	RGLRVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQOAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRNRALNL
25	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPOPSSFA	QTIPOYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGRYSRIRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
	401	FVPQLQKDEH	GSYLKPVGTN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
30	451	TAAGRQSGN	TYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRINPDV	PERSEKLFET	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAKAR	AVANSRQKAY	AVADIMARYR	FNPRTELSLN	VDNLFNKHYR
	601	TQPRHSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACC GC	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCGGGC	TGCCCATGAC
	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GGCGACCGGC
	301	ACCAGCCGCC	AGATTACGG	CTCGACCGCG	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAACTACCA	AATCAACGGC	ATCCCGTTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGGCGT	GGCGGGGCTG	CCGACCGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCGCG	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGCGG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAACT	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
50	751	GTCCACGCAG	GATGGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCTG	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCG	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	GACGGCTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCgatatcc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAATATC	CgcctGTTTCG	GCCGCGAGCA	CGATTTAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTTCC
	1201	AACGCCATTG	CCAACGCCTA	CGAATTTTCC	CGCACGGGCG	CCTATCCGCA
60	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTGCTCG
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCGC	GCAGGCAGCT	ACAACAGCCG

-384-

1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTCACC CCCTACACAG
 1451 GCATCGTGTT CGATCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC
 1501 AGCCTGTTTCG TCCCGCAATT GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGGAAGCCGA CATCAAAGGC GAATGGCTTG
 5 1601 AAGGGCGTCT GAACGCATCC GCCGCCGTGT ACCGCCGCCG TAAAAACAAC
 1651 CTCGCCACCG CAGCAGGACG CGACCAGAGC GGCAACACCT ACTATCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGCGGCCGCA
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGCT ACAGCCAAAG CAAACCCCGC
 1801 GACCAAGACG CGAGCCGCCT GAACCCCGAC AGCGTACCCG AACCGAGCTT
 10 1851 CAAACTCTTC ACCGCCTACC ACTTAGCCCC CGAAGCCCCC AGCGGCCGGA
 1901 CCATcgTGC GGGTGTGCGC CGGCAGGGCG AAACCCACAC CGACCCAGCC
 1951 GCGCTCCGCA TCCCCAACCC CGCCGCCAAA GCCCGCGCCG TCGCCAACAG
 2001 CCGCCAGAAA GCCTACGCCG TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCAC CGAAGTGTCT CTGAACGTGG ACAACCTGTT CAACAAACAC
 15 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCGACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVSDPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PFGLPMTLRE IPQSVSVITS QQMRDQNIKT LDRALLQATG
 20 101 TSRQIYGS DR AGYNLFARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL PDGTGEPSAT VNLVRKHPT R KPLFEVRAEA GNRKHFGLGA
 201 DVSGSLNAEG TLRGRLVSTF GRGDSWRQLE RSRDAELYGI LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWSNSRNRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH
 25 351 GYWHADPRTH SASMSLTGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPQSS FAQTIPQYDT RRQIGGYLAT RFRADNLSL
 451 ILGGYSRYR AGSYNSRTQG MTYVSANRFT PYTGIVFDLT GNLISLYGSYS
 501 SLFVPQLQKD EHGSYLKPVT GNNLEADIKG EWLEGRNLNAS AAVYRARKNN
 551 LATAAGRDQS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKPR
 601 DQDGSRLNPD SVPERSFKLF TAYHLAPEAP SGRTIGAGVR RQGETHTDPA
 651 ALRIPNPAK ARAVANSRQK AYAVADIMAR YRFNPRTELS LNVNDLNFKNH
 701 YRTQPRHSY GALRTVNAAF TYRFK*

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35	orf23-1.pep	10	20	30	40	50	60
	orf23ng-1	10	20	30	40	50	60
40	orf23-1.pep	70	80	90	100	110	120
	orf23ng-1	70	80	90	100	110	120
45	orf23-1.pep	130	140	150	160	170	180
	orf23ng-1	130	140	150	160	170	180
50	orf23-1.pep	190	200	210	220	230	240
	orf23ng-1	190	200	210	220	230	240
55	orf23-1.pep	250	260	270	280	290	300
	orf23ng-1	250	260	270	280	290	300
60	orf23-1.pep	310	320	330	340	350	360
	orf23ng-1	310	320	330	340	350	360

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYRLFGREHDLIAGINGKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
	orf23ng-1	SASMSLTGKYRLFGREHDLIAGINGKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPSS					
		370	380	390	400	410	420
10	orf23-1.pep	FAQTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
	orf23ng-1	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYSTRAGSYNSRTQGMTYVSANRFT					
		430	440	450	460	470	480
15	orf23-1.pep	PYTGIVFDLTGNLSLYGSYSSLFVPSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLNAS					
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVPSQKDEHGSYLKPVTGNNLEADIKGEWLEGRNLNAS					
		490	500	510	520	530	540
20	orf23-1.pep	AAVYRARKNNLATAAGRDPSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR					
	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKPR					
25		550	560	570	580	590	600
30	orf23-1.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDTPATLRIPNPAK					
	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAK					
		610	620	630	640	650	660
35	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF					
	orf23ng-1	ARAVANSRQKAYAVADIMARYRFNPRTLSLNVNLFNKHYRTQPDHRSYGALRTVNAAF					
		670	680	690	700	710	720
40	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrucic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDQNIKTLDRL 95 T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSTAGTKMQMTQRDIPQSVTIVSQRMEDQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNYLFARGSRANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSRLALY---YSRGFQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEP SATVNLRKHPTRKPLF-EVRAEAGNRKHFG LGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSP 214
70	Query: 207 NAEGTLRGLRVSTFGRGDSWRQLERSDAELYGILEYDIAPQTRVHAGMDYQAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFEFGIVDADLGDLTTL SAGYEYQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNRLNLFAGIEHRFNQDWKLKAEYDYTRSR 326

+++ G + ++ + A +W+ + +F ++ +F W+ ++
 Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLF 374
 F + Y A V D ++ PG+ W++ R A + G Y LFG
 Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394
 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
 Sbjct: 395 RQHNLMEFG-GSYSKQNNRYFSSWANIFPDEIGSFYNFN--GNFPQTDWSPQSLAQDDTTH 451
 Query: 433 QIGGYLATRFRAADNLSLILGGRYSRYRAGSYNSRTQGMTY-VSANRETPYTGIVFDXXX 491
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504
 Query: 492 XXXXXXXXXXXXFPVQLQKDEHGSYLKPVGTGNLEADIKGEWLEGRNLSAAVYRARKNNL 551
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
 Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564
 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRIPEWQIQAGYSQSKPRDQDGSRLN 608
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTYSKGVFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624
 Query: 609 PDSVPERSEKLFETAYHLAPEAPSGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668
 P ++P + K+FT+Y L P P T+G GV Q +TD P RA
 Sbjct: 625 P-NLERTTVKMETSURL-PVMPE-LTVGGGVNWQNRVYTDTV-----TPYGTFRFRA----E 672
 Query: 669 QKAYAVADIMARYRFPNRTLSLNVNLFNKHYRTQPDHR-SYGALRTVNAAFTYRF 724
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F
 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
 151 AGCGTCAGcA CGCCTGCTTC GCGGgcGgCa ATCATACCTT CGTCTTCGGA
 201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
 251 TGCCGCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
 301 CCGTGCCTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
 351 TnAGTCGCCC ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTQ QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTCC
5   151  AGCGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
201  AACGGGGATA AACGCGCCAC TCAAACCCCC GACCGCGCTG GAAGCCATCA
251  TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301  CCGTGCCTAC CGCAGACGCT CAAGCCCAT TCTTCAAGAA TGCCTGCCAC
351  TGAGTCGCGG ACGGCGGGGG TCGGCGCCAG CGACAAGTCG AGAATACCAA
10  401  ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG
451  CGGGTAATTT TGAAAGCAGT TTCTTCACT ACTTCCGCAA CTTCGGTCAA
501  TGTCGTTGCA TCTGAATTTT CCAACGCGGC TTTTACGACA CCTGGGCCGG
551  ATACGCGGAC ATTGATAACG GCATCCGCTT CGCCCCGAAC ATGAAACGCG
601  CCCGCCATAA ACGGGTTGTC TTCCACGCGG TTGCAGAACA CGACAATTTT
15  651  AGCGCAGCCG AAACCTTCGG GCGTGATTTC CGCCGTGCGT TTGACGGTTT
701  CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTACTGCCG
751  ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTCGGGAAT
801  GGAGCGGATT AACACCTCAT CCGAAGCGCA CATCCCTTTT TGCACCAACG
851  CGGAAAAACC GCCGATAAAA GACACACCGA TGGCTTTGGC AGCTTTATCC
20  901  AAAGTTTGCG CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSSPT
25  151  RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
201  PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLP
251  ILMELHTISV VFASGMERI NTSSEGDIPF CTNAEKPPIK DTPMALAALS
301  KVCATLT*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N.*

meningitidis:

```

35  orf24a.pep      10      20      30      40      50      60
      MRTAVVLLLIMPMAASSAMMPMVCAVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
      |||||:|||||:|||||
      orf24        10      20      30      40      50      60
      MRTAVVLLLIMPMAASSAMMPMVCAVSPGTAIISKPTEQTAVMASSLSNVSTPASAAA

40  orf24a.pep      70      80      90      100     110     120
      IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP
      |||||:|||||:|||||:|||||
      orf24        70      80      90      100     110     120
      IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAAVVPCVPQTLKPISSRMRATESP

45  orf24a.pep      130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
      |||||:|||||:|||||:|||||
      orf24        130     140     150     160     170     180
      TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT

50  orf24a.pep      190     200     210     220     230     240
      PGPDTPTLITASASPEPXNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
      |||||:|||||:|||||:|||||
      orf24        190     200     210     220     230     240
      PGPDTPTLITASASPEPXNAPAIINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA

55  orf24a.pep      250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
      |||||:|||||:|||||:|||||
      orf24        250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS

```

```

orf24a.pep  KVCATLTx
|||||||
orf24       KVCATLTx

```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

1  ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
151 AACGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
10  201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACCGCGCTC GAAGCCATCA
251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
301 CCGTGCGTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGC CGCGCCAC
351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCCACG
15  451 CGGGTAATTT TGAAGGCGGT TTTCTTCACA ACTTCGGCAA CTTCGGTCAA
501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
601 CCCGCCATAN ACGGTTTGTC TTCCNCCGCG TTGCAGACA CGACGATTTT
651 GGCGCAGCCG AAACCTTCTA GTGTGATTTC ANCCGTGCGT TTGATGGTTT
20  701 CGCCGCGCAG TCTGACCGCG TCCATATTGA TACGGCGCGC CGTACTGCCG
751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
801 GGAACGGATN AACACCTCGT CAGAAGGCGA CATACCTTTT TGCACCAGCG
851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
901 AAAGTTTGCG CCACGCTGAC GTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
51  NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASREMSST
151 RVILKAVFFT TSATSVNVVA SEFSNAFFT PGPDTPTLIT ASASPEP*NA
30  201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
251 ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
301 KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS NVSTPASAAA
          |||
orf24-1      MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS NVSTPASAAA
          |||
40      70      80      90      100     110     120
orf24a.pep  IIPSSSXTGI NAPLKPPTALEAIMPPFFTA SFSNAKAAVV PCVPQTLKPI SSRMRATESP
          |||
orf24-1      IIPSSSXTGI NAPLKPPTALEAIMPPFFTA SFSNAKAAVV PCVPQTLKPI SSRMRATESP
          |||
45      70      80      90      100     110     120
orf24a.pep  TAGVGASDKS RIPNGIFSIF EASRPMSPT RVILKAVFFTT SATSVNVVASEFSNAFFT
          |||
50  orf24-1      TAGVGASDKS RIPNGIFSIF EASRPMSPT RVILKAVFFTT SATSVNVVASEFSNAFFT
          |||
          130     140     150     160     170     180
orf24a.pep  TAGVGASDKS RIPNGIFSIF EASRPMSPT RVILKAVFFTT SATSVNVVASEFSNAFFT
          |||
50  orf24-1      TAGVGASDKS RIPNGIFSIF EASRPMSPT RVILKAVFFTT SATSVNVVASEFSNAFFT
          |||
          130     140     150     160     170     180
orf24a.pep  PGPDTPTLIT ASASPEPXNAPAIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
          |||
55  orf24-1      PGPDTPTLIT ASASPEPXNAPAIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
          |||
          190     200     210     220     230     240
orf24a.pep  PGPDTPTLIT ASASPEPXNAPAIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
          |||
55  orf24-1      PGPDTPTLIT ASASPEPXNAPAIXGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
          |||
          190     200     210     220     230     240
orf24a.pep  SILIPARVLP ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
          |||
60  orf24-1      SILIPARVLP ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
          |||
          250     260     270     280     290     300

```

orf24a.pep KVCATLT
 5 orf24-1 KVCATLT

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

10 orf24.pep MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTQTAVMASSLSVSTPASAAA 60
 orf24ng MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIMSKPTQTAVMASSLSVNTPASAAA 60
 15 orf24.pep IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPIXSRMRATXSP 120
 orf24ng IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRAATESP 120
 orf24.pep TG 122
 20 orf24ng TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVRLTASEFSSAALT 180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

1 ATGCGCACGG CGGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GCGGATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
 101 TCATGTCCAA ACCAACGGAG CAGACGGCGG TCATGGCTTC GAGTTTGTCC
 25 151 AGCGTCAACA CGCTGCCTC GCGGCGGCA ATCATACCTT CGTCTTCGGA
 201 AACGGGGATA AACGCGCCGC TCAAACCGCC GACCGCGCTG GAAGCCATCA
 251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
 301 CCGTGCGTAC CGCAGACGCT CAAGCCCATT TCTTCAAGAA TGCGCGCCAC
 351 CGAGTCGCGG ACGCGGGGGG TCGGTGCCAG CGACAAATCG AGAATGCCGA
 401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GACCGATGAG TTCGCCCACG
 451 CGGGTGATTT TGAAGCGGT TTTCTTCACG ACTTCGGCGA CCTCGGTCAG
 501 GCTGACCGCG TCCGAATTTT CCAGCGCGGC TTTGACCACG CCTGGACCGG
 551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCCGAGCC GTGGAACGCA
 601 CCCGCCATAA ACGGATTGTC TTCCACCGCG TTGCAGAACA CGACGATTTT
 35 651 GCGCGACGGG AAACCTTCGG GTGTGATTTC AGCCGTGCGT TTGATGGTTT
 701 CGCCTGCCAG CTTGACCGCA TCCATATTGA TACCGGCACG CGTCTGCCG
 751 ATATTGATGG AGCTGCACAC GATATCGGTA GTTTTCATCG CTTCGGGAAC
 801 GGAACGATC AACACCTCAT CCGAAGGCGA CATACCTTTT TGCACGAGCG
 851 CGGAAAAGCC GCCGATAAAG GACACGCCGA TGGCTTTGGC TGCCTTGTCC
 40 901 AAAGTCTGCG CCACGCTGAC ATAA

This encodes a protein having amino acid sequence <SEQ ID 680>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIMSKPTE QTAVMASSLS
 51 SVNTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAAVV
 101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RMPNGIFSIF EASREMSPT
 45 151 RVILKAVFFT TSATSVRLTA SEFSSAALT PGPDTPTLIT ASASPEPUNA
 201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LMVSPASLTA SILIPARVLP
 251 ILMELHTISV VFIASGTERI NTSSEGDIPF CTSAEKPPIK DTPMALAALS
 301 KVCATLT*

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50 orf24-1.pep 10 20 30 40 50 60
 MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIISKPTQTAVMASSLSVSTPASAAA
 orf24ng MRTAVVLLLIMPMAASSAMMPMVCAGVSPGTAIMSKPTQTAVMASSLSVNTPASAAA
 55 10 20 30 40 50 60
 orf24-1.pep 70 80 90 100 110 120
 IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRAATESP
 orf24ng IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVPCVPQTLKPISSRMRAATESP
 60 70 80 90 100 110 120

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT					
5	orf24ng	TAGVGASDKSRMPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVRLTASEFSSAALT					
		130	140	150	160	170	180
	orf24-1.pep	190	200	210	220	230	240
10	orf24ng	PGPDTPTLTITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
		190	200	210	220	230	240
	orf24-1.pep	250	260	270	280	290	300
15	orf24ng	SILIPARVLPILMELHTISVVFIA SGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
		250	260	270	280	290	300
20	orf24-1.pep	KVCATLTX					
	orf24ng	KVCATLTX					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-
underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein,
it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could
be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

```

1  ..ACCGACGTGC AAAAAGAGTT GGTCGGCGAA CAACGCAAGT GGGCGCAGGA
51  AAAAATCAGC AACTGCCGAC AAGCCGCCGC GCAGGCAGAC CGGCAGGAAT
101 ACGCCGAATA CCTCAAGCTG CAATGCGACA CGCGGATGAC GCGCGAACGG
151 ATACAGTATC TTCGCGGCTA TTCCATCGAT TAG

```

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

```

1  ..TDVQKELVGE QRKWAQEKIS NCRQAAAQAD RQEYAEYLKL QCDTRMTRER
51  IQYLRGYSID *

```

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

```

1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
40 101 TGCAAGGCAT ACGCGGCAAT ATTCAAGAAA CGCTCAGCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACGG CAGGCAGTTT GTCGATGCCG ACAAATTAT
201 CGCCGCCGCG TACGTTTGG CGTTTCTTT GGAACACGCT TCGGAAACGC
251 AGGAAGGCGG GCGCACGTTT TGTATCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CAAGGCAAAC AGCCCCTGT TGTACGGGGA
45 351 AACTGCTTTG TCGGATATTG TCGGCAGAA GACGGGCGGC AATGTCGAGT
401 TTAAAGACGG CGTATTGACG GCAGCCGTCC GCTTCCTGCC CGTCAAAGAC
451 GGTACAGACGG CATTGTGCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
501 GTCTGCCGCG CTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT GAAAAAGAA GACGCGGTCA GGATTTGAG CGGAAAAGCC
50 601 CGTGAAGAAG AACCGTCCAA ACCACGCCCC GAAGACATTT TGGAACACAA
651 TGCCGCCGCG GCGATGCGG GCGTACCCCA AGCCGAGAA GCGCGCCCG
701 AACCGGAAAT CCTGCATCCT GACGACGGCG AGCGTGCCGA TACCGTTACC
751 GTATACGGG GCGAAGTGGA AGAGGCGCGC GTACAAAACC AGCGTGCGGA
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
55 851 AGTTGGTCGG CGAACAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
1001 GCTATTCCAT CGATTAG

```

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQGIRGN IQETLTQEAR
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
5  151 GQTAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEPEPKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAQAADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N. meningitidis*:

```

15 orf25.pep                      10      20      30
                                TDVQKELVGEQRKWAQEKISNCRQAAAQAD
                                |||||
orf25a      VTVSRGEVEEARVQNQRAESEITKLWGGLDLDVQKELVGEQRKWAQEKISNCRQAAAQAD
            250      260      270      280      290      300

20 orf25.pep                      40      50      60
                                RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
                                |||||
orf25a      RQEYAEYLKLCQDTRMTRERIQYLRGYSIDX
            310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25 1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTGT
51  CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCAGCA GGAAGCGCGT
151 TCTTTTCGCG CCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
201 CGCCGCCGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
30 251 AGGAAGGCGG GCGCACGTTT TGTNTCGCCG ATTTGAACAT TACCGTGCCG
301 TCTGAAACGC TTGCCGATGC CAAGGCAAC AGCCCCCTGC TGTACGGGGA
351 AACCGCTTTG TCGGATATTG TGCGGCAGAA GACGGCGCGC AATGTGAGT
401 TTAAGACGCG CGTATTGACG GCAGCGCTCC GCTTCTTACC CGTCAAGAC
451 GGTCAAGANG CATTTGTCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
35 501 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT AAAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
601 CGTGAANAAG AACCGTCCAA ANCCNNGCCC GAAGACATT TGGAAACATAA
651 TGCCGCCGGA GGGGATGCAG ACGTACCCCA AGCCGAGAGG GACGCGCCCG
701 AACCGGAAAT CCTGATCCTT GACGACGGCG AGCGTGCCGA TACCGTTACC
40 751 GTATCACGGG GCGAAGTGGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
801 ATCCGAAATT ACCAACTTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCCG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
45 1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
51 SFAREDXXQF VDADKIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
101 SETLADAKAN SPLLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
50 151 GXQAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRIXSXXA
201 REXEPSKXXP EDILEHNAAG GDADVQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGGLDLD VQKELVGEXR KWAQEKISNC
301 RQAAQAADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

55 orf25a.pep      10      20      30      40      50      60
                                MYRKLIALPFALLLAACGRE EPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
                                |||||
orf25-1          MYRKLIALPFALLLAACGRE EPPKALECANPAVLQIRGNIQETLTQEARSFAREDGRQF

```


-392-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf25a.pep	VDADXIIAAAXXXXSLEHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL					
	orf25-1	VDADKIIAAAYGLAFSLEHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL					
		70	80	90	100	110	120
10	orf25a.pep	130	140	150	160	170	180
	orf25-1	130	140	150	160	170	180
15	orf25a.pep	190	200	210	220	230	240
	orf25-1	190	200	210	220	230	240
20	orf25a.pep	250	260	270	280	290	300
	orf25-1	250	260	270	280	290	300
25	orf25a.pep	310	320	330	339		
	orf25-1	310	320	330	339		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLD TDVQKELVGEQRKWAQEKISNCRQAAAQAD	308
40	orf25.pep	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	60
	orf25ng	RQEYAEYLKLCQDTRMTRERIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACTCATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTCAGGAAA	CGCTCACGCA	GGAAGCGCGT
	151	TCTTTTCGCG	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTCTTTT	GGAACACGCT	TCGGAACGCG
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
50	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTTATCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
55	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTGTAG	CGGCAAGGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCC	GAAGACATTT	TGGAACACAA
	651	TGCCCGCCGC	GCGATGCGG	GCGTACCCCA	AGCCGAGAA	GCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
60	801	ATCCGAAATT	ACCAAACCTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCGG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
 101 SETLADAEAN SPLLYGETSL ADIVQQKTGG NVEFKDGLT AAVRFLPAKD
 151 ARTAFIDNTV GMATQTLASAA LLPYGVKSIV MIDGKAVTKE DAVRVLSGKA
 201 REEEPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT
 251 VSRGEVEEAR VQNQRAESEI TKLWGGLD TDVQKELVGEQR KWAQEKISNC
 301 RQAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330	339		

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described
 50 above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and
 55 that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
      51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
     101  GCATCGGTAT TCTGGWysGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
     151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
     201  CGsyGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTT CkGATACTTT
     251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....
10      //
      851  .....AC TTCGCTGGTA
      901  TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
      951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
     1001  TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
     151  1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
     1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGTGTC GCCAGCGTGA
     1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
     1201  ATTGCCGCGG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
     1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
     1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
     1351  GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
     1401  CGCATCGGGG TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTGGGGCT
     1451  TTGGCAGCAC AGGCATTGTA TTGGCGGTGC TGATTTTCTT GTTGAAAGAT
     1501  AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1  MQLIDYSHSF FSVVPEFLAL ALAVITRRVL LSLGIGILXX VAFLVGGNPV
     51  DGLTHLKDMV VGLAWSDXDW SLGKPKILVF XILLGIFTSI LTYSGSN...
      //
     251  .....TSLV
     301  FGGTCGVFAV VLCTLGTIKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
     351  VGEMHTGDYL STLVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
     401  IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARCNIH
     451  DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTGIV LAVLIFLLKD
     501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1  ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
     51  TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
    101  GCATCGGTAT TCTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
    151  GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
    201  CGGCGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTT CTGATACTTT
    251  TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
    301  GCCGACTGGG CAAACCGGCA CATTAATAAC CGGCGCGGCG CGAAAATGCT
    351  GACCGCTGCG CTCGTGTTTC TAACCTTTAT CGACGACTAT TTCCACAGTC
    401  TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
    451  CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
    501  GCTGATGCCC GTTTCAGGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
    551  GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCCGAT GGGGACGTTT
    601  GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
    651  GTTCGTCGTC GCATGGTTT CCTTCGACAT CGGCTCGATG GCACGTTTCG
    701  AACAAAGCCG GTTGAACGAA GCCCAGCATG AAATGCCGT TTCAGACGCT
    751  ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTTGG CCTTAATCGC
    801  CTCAACGGTT TCCGCCATGA TCTACACCGG CGCGCAGGCA AGCGAAACCT
    851  TCAGCATTTT GGGGCGATTT GAAAACACGG ACGTAAACAC TTCGCTGGTA
    901  TTCGGCGGCA CTTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
    951  GATTAAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
   1001  TGTTCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
   1051  GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
   1101  CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGTGTC GCCAGCGTGA
   1151  TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
   1201  ATTGCCGCGG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
   1251  TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
   1301  TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

-395-

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
 1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
 1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
 1501 AAAAAACGCG CCAACGCCTG A

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMV VGLAWSGDGW SLGKPKILVF LILGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHS LAVGAIA RPVTDKFKVS
 151 RTKLAYILDS TAAFCVCLMP VSSWGASIIA TLAGLLVYK ITEYTPMGTF
 10 201 VAMSLMNYA L FALIMVFV AWFSFDIGSM ARFEQAALNE AHDETAVSDA
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTLGITKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
 351 VGMHTGDYL STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRANA*

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

Orf26 1 MQLIDYSHSFSSVPPFLALALAVITRRVXXXXXXXXXXVAFLVGGNPVDGLTHLKDMV 60
 M+LID+S S +S+VP LA+ LA+ TRRV L +L V
 HI1586 14 MELIDFSSSVSIVPALLAIIAIAIATRRVLVSLSAGIIIGSLMLSDWQIGSAFNYLVKNV 73
 25 Orf26 61 VGLAWSDXDWSLGKPKILVFXILGIFTSLTYSGSN 97
 V L ++D + + I++F +LLG+ T+LLT SGSN
 HI1586 74 VSLVYADGEIN-SNMNIVLFLLLGLTALLTVSGSN 109
 //
 30 Orf26 86 IFTSLLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPKAVWQGA KSMFGXXX 141
 +F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
 HI1586 299 VFSVLGTFENTVVGTSVLVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGKSMGSAIAI 358
 35 Orf26 142 XXXXXXSTVVGEMHTGDYLSTLVAGNIHPGFLPVILFLLASVMAFATGTSGWTFGIMLP 201
 + +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSGWTFGIMLP
 HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSGWTFGIMLP 418
 40 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDH CSPISDTTILSSTGARCNIHIDVTSQXXX 261
 IAAAMA P L++PC+SAVMAGAVCGDH CSP+SDTTILSSTGA+CNHIDHVT+Q
 HI1586 419 IAAAMAANAAPELLPLCLSAVMAGAVCGDH CSPVSDTTILSSTGAKCNHIDHVTQLPYA 478
 Orf26 262 XXXXXXXXXXXXXXXXSALLGFGTTGIVLAVLIFLLKDK 302
 S L GF T + L V+IF +K +
 45 HI1586 479 ATVATATSIGYIVVGFTYSLAGFAATAVSLIVIIFAVKKR 519

Homology with a predicted ORF from *N. meningitidis* (strain A).

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

meningitidis:

50 orf26.pep 10 20 30 40 50 60
 MQLIDYSHSFSSVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV
 orf26a MQLIDYSHSFSSVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
 55 orf26.pep 70 80 90 99
 VGLAWSDXDWSLGKPKILVFXILGIFTSLTYSGSNXX-----
 orf26a VGLAWSGDGWSLGKPKILVFLILGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
 60 70 80 90 100 110 120

1	ATGCAGCTGA	TGCACATATC	ACATCAATCTT	TTCTTCGGTTG	TGCCACCCCTT
51	TTTGGCACTG	TCGACTTGGC	TCATTACCCG	CCGCGTACTG	CTGCTCTTAG
101	GCATCGGTAT	TCTGGTGGC	GTTGCCCTTT	TGGTCGGCGG	CAACCCCGTC
151	GACGCTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGTCAGA
201	CGGCGATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTC	CTGACTACTT
251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGGACCAA	TCAGGCGTTT
301	GCCGACTGGG	CAAAAACGGCA	CATTAAAAAC	CGGCGCGCGG	CGAAAAATGCT
351	GACGCGCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
401	TGCGCGTCGG	TGCGNNTGCC	CGCCCCGTTA	CCGACAAAGT	TAAAGTTTCC
451	CGCGCCAAAC	TCGCCCTACAT	CCTCGACTCC	ACTGCCGCGC	CTATGTGCGT
501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTCCCG
551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
601	GTGCGCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGGCC	TGATTATGGT
651	GTTCTGTCGT	GCATGGTTCT	CCTTCGACAT	CGGCTCGATG	GCACGTTTCG
701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTTCAGACGGC
751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCCGTTTGGG	CTTCAATCGC
801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
851	TCAGCATTTT	GGGTGCATTT	GAAAATACGG	ACGTGAACAC	TTTCGCTGGTA
901	TTGCGCGGCA	CTTGCGGCGT	GCTTGCCGTC	GTCTCTGCA	CGCTCGGCAC
951	GATTAATAATC	GCCGATTATC	CCAAAGCCGT	TTGGCAGGTA	GCGAAATCCA
1001	TGTTCCGGCG	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
1051	GTCCGCGGAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTCGCGGCGAA
1101	CATCCATCCC	GGCTTCTGTN	CCGTATCCTT	TTTCTGTCTC	GGCAGCGTGA
1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCGGCAT	CATGCTGCCG
1201	ATTGCCGCGC	CCATGGCGGT	CAAAGTCGAT	CCCTCAGCAT	TTATCCGCTA
1251	TATGTCCGCG	GTCATGGCGG	GGGCGGTATG	CGGCGACCCAC	TGCTCGGCCA
1301	TTTCCGACAC	GACCATCCTG	TCGTCCACCG	CGGCGGCGTG	CAACCATCA

-397-

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
 1501 AAAAAACGCG CCAACGCCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVGGNPV
 51 DGLTHLKDMV VGLAWSGDW SLGKPKXLVF LILLGIFTS LTYSGSNQAF
 101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDDY FHSLAVGAXA RPYTDKFKVS
 151 RAKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
 10 201 VAMSLMNYIA LFALIMVFVW AWFSFDIGSM ARFEQAALNE AHDETAVSDG
 251 SWGRVYALII PVLALIASTV SAMIYTGQAQ SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTLGITIKI ADYPAKVWQG AKSMFGAIAI LILAWLISTV
 351 VGEMHTGDYL STLAVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVD PSLIIPCMA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 15 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
 501 KKRANA*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV					
	orf26-1	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV					
		10	20	30	40	50	60
25	orf26a.pep	70	80	90	100	110	120
	orf26-1	VGLAWSGDWVSLGKPKXLVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRRGAKMLTAC					
		70	80	90	100	110	120
30	orf26a.pep	130	140	150	160	170	180
	orf26-1	LVFVTFIDDYFHSLAVGAXARPYTDKFKVSRKALAYILDS TAAPMCVLMFVSSWGASIIA					
		130	140	150	160	170	180
35	orf26a.pep	190	200	210	220	230	240
	orf26-1	TLAGLLVITYKITEYTPMGTFVAMSLMNYIALFALIMVFVWAWFSFDIGSMARFEQAALNE					
40		190	200	210	220	230	240
45	orf26a.pep	250	260	270	280	290	300
	orf26-1	AHDETAVSDGSWGRVYALII PVLALIASTVSAMIYTGQAQSETFSILGAFENTDVNTSLV					
		250	260	270	280	290	300
50	orf26a.pep	310	320	330	340	350	360
	orf26-1	FGGTCGVLAVVLCTLGITIKIADYPAKVWQGA KSMFGAIAI LILAWLISTVVGEMHTGDYL					
		310	320	330	340	350	360
55	orf26a.pep	370	380	390	400	410	420
	orf26-1	STLVAGNIHPGFLXVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVDPSLIIPCMA					
		370	380	390	400	410	420
60	orf26a.pep	430	440	450	460	470	480
	orf26-1	VMAGAVCGDHCSPISTTILSSTGARCNIHDHVTSQLPYALTVA AAAASGYLALGLTKSA					
		430	440	450	460	470	480
65	orf26a.pep	490	500	LLGFGXTGIVLAVLIFLLKDKKRANAX			

-398-

or26-1 |||||:|||||:|||||:|||||:|||||
 LLGFGTTGIVLAVLIFLLKDKKRANAX
 490 500

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	or26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFLVGGNPVDGLTHLKDMV	60
10	or26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVGGNPVDGLTHLKDMV	60
	or26.pep	VGLAWSDXDWSLGKPKILVFXILLGIFTSLLTYSGSN	97
15	or26ng	VGLAWADGDWSLGKPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	or26.pep	TSLVFGGTCGVFAVVLCTLGTIKTADYPKA	326
20	or26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTCGVFAVVLCTFGTIKTADYPKA	326
	or26.pep	VWQGAISMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF	386
25	or26ng	VWQGAISMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF	386
	or26.pep	ATGTSWGTFGIMLPAAAAVAVKEPALIIPCMSAVMAGAVCGDHCSPISTTILSSTGAR	446
	or26ng	ATGTSWGTFGIMLPAAAAVAVKEPALIIPCMSAVMAGAVCGDHCSPISTTILSSTGAR	446
30	or26.pep	CNHIDHVTSQLPYALTAAAAAGSYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	or26ng	CNHIDHVTSQLPYALTAAAAAGSYLALGLTKSALLGFGTTGIVLAVLIFLLKDKKRAADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

	1	ATGCAGCTGA	TTGACTATTC	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
35	51	TTTGGCACTG	GCACCTGCGG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TTTGGTGGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGGCTTG	CTGGGGCAGA
	201	CGGGGATTGG	TCGCTGGGCA	AACCAAAAAT	CTTGGTTTTC	CTGATACTTT
40	251	TGGGCATTTT	CACTTCACTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGTGCGGCG	CGAAAAATGCT
	351	GACCGCCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGCC
	401	TGCGCCCTCG	TGCGATTGCC	CGCCCGGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCTACAT	CCTCGACTCC	ACTGCCTCGC	CCATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
45	551	GATTGCTCGT	TACCTACAAA	ATTACCGAAT	ACACGCCGAT	GGGGACGTTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCG	CTGTTTGCCC	TGATTATGGT
	651	ATTCTGTCGT	GCATGGTTCT	CCTTCGACAT	CGGCTCGAtg	gCGCGTTTCG
	701	AACAGGCTGC	GTTGAACGAA	gcccaggacg	aaaccgcccgc	tTCAGACgCT
	751	ACCAAAGGTC	GTGTTTACGC	ATTGATTATT	CCCGTTTGGG	CCTTAATCGC
50	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	CGCGCAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGGGCATTT	GAAAATACCG	ACGTAAACAC	TTCGCTGGTA
	901	TTCGGCGGCA	CTTGCGGCGT	GCTTGCGGTC	GTCCTCTGCA	CGTTCGGCAC
	951	GATTAAAACC	GCCGATTATC	CCAAAGCCGT	GTGGCAGGGT	GCGAAATCCA
	1001	TGTTCCGGCG	AATCGCCATT	TTAATCCTCG	CCTGGCTCAT	CAGTACGGTT
55	1051	GTCGGCGAAA	TGCACACGGG	CGACTACCTC	TCCACGCTGG	TGCGGGGCAA
	1101	CATCCATCCC	GGCTTCTCTG	CCGTATCCTT	CTTCTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCGCGAT	TATGCTGCCG
	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAA	CCCGCGCTGA	TTAtcccGTG
	1251	TATGTCCGCA	GTAATGGCGG	GGGCGGTATG	CGGCGACCAC	TGTTCCGCCA
60	1301	TCTCCGACAC	GACCATCCTG	TCGTCCACCG	GCGCGCGCTG	CAACCACATC
	1351	GACCAAGTTA	CCTCGCAACT	GCCTTATGCC	CTGACGGTTG	CCGCGCGCGC
	1401	CGCATCGGGC	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTTGGGCT
	1451	TTGGCACGAC	CGGTATTGTA	TTGGCGGTGC	TGATTTTCTT	GTGAAAGAT
	1501	AAAAACCGG	CCGACGTTTG	A		

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

-399-

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGGNPV
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDY FHSLAVGAIA RPVTDKFKVS
 151 RAKLAYILDS TASEMCMVLP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
 5 201 VAMSLMNYA LFALIMVFVW AWFSDIGSM ARFEQAALNE AQDETAASDA
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTFGTIKT ADYPAVWQG AKSMFGAIAI LILAWLISTV
 351 VGEHMTGDYL STLAVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 10 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRADV*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

15	orf26-1.pep	10	20	30	40	50	60
	orf26ng	10	20	30	40	50	60
20	orf26-1.pep	70	80	90	100	110	120
	orf26ng	70	80	90	100	110	120
25	orf26-1.pep	130	140	150	160	170	180
	orf26ng	130	140	150	160	170	180
30	orf26-1.pep	190	200	210	220	230	240
	orf26ng	190	200	210	220	230	240
35	orf26-1.pep	250	260	270	280	290	300
	orf26ng	250	260	270	280	290	300
40	orf26-1.pep	310	320	330	340	350	360
	orf26ng	310	320	330	340	350	360
45	orf26-1.pep	370	380	390	400	410	420
	orf26ng	370	380	390	400	410	420
50	orf26-1.pep	430	440	450	460	470	480
	orf26ng	430	440	450	460	470	480
55	orf26-1.pep	490	500				
	orf26ng	490	500				

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037
 hypothetical
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
 Score = 538 bits (1370), Expect = e-152
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5 Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXXAFIVGGNPVDGLTHLKDMMV 60
 M+LID+S S +S+VP LA+ LA+ TRR L +L V

10 Sbjct: 14 MELIDFSSSVWSIVPALLAIILAIATRRVLVSLSAGIIGSLMLSDWQIGSAFNVLVKNV 73

Query: 61 VGLAWADGWSLKGPKILVFLILLGIFTSLTTYSGSNQAFADWAKRHIKRCGAKMLTAC 120
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A

15 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGLVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDDFHSLAVGAIARPVTDFKVSRAKLAYILDSTASPMCVLMPVSSWGASIIA 180
 LVFVTFIDDFHSLAVGAIARPVTD+FKVSRAKLAYILDSTA+PMCV+MPVSSWGA II

20 Sbjct: 133 LVFVTFIDDFHSLAVGAIARPVTDREFKVSRAKLAYILDSTAAPMCVMMMPVSSWGAYIIT 192

Query: 181 TLAGLLVTYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE 240
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL

25 Sbjct: 193 LIGLLATYSITEYTPIGAFVAMSSMNFYAFISIIMVFFVAYFSFDIASMVRHEKLALKN 252

Query: 241 AQDETAASDATKGRVYALIIIPVLALIASTVSAMIYTGAAQ----SETFSILGAFENTDVN 296
 +D+ TKG+V LI+P+L LI +TVS MIYTGAA+ + FS+LG FENT V

30 Sbjct: 253 TEDQLEEETGKQVRNLILPILVLIATVSMIYTGAEALADGKVFVSLGTFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQAKSMFGXXXXXXXXXXSTVVGEM 354
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M

35 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALI 414
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLPAAAMA P L+

40 Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSWGTFGIMLPAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISDITILSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXXXXX 474
 +PC+SAVMAGAVCGDHCSP+SDTILSSTGA+CNHIDHVT+Q

45 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSDTILSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
 S L GF T + L V+IF +K +

Sbjct: 493 GFTYGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

50 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51 CGATGAGCCT GCCAAAATTC TGAATGCGGA TGAAAGCGGC CGATTACTCT
 101 CGGAACCTGTC TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
 201 GGTACAGAAA ACCCAGTGGG ATAAGGATGG TTATTAAATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

55 1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HQRNGVVLEW
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGyli EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

60 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

5
10
201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT
401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTACCGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GGCGTATGGA AGCAATGGTA
501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
551 CCAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

15
1 MKKLSRIVFS TVLLGFS AAL PAQTYSVYFN QNGKLTATMS SAAYIROYSV
51 VAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
101 KMAGGFSK GK PDGEWVNWYP NGKKS AVMPY KNGLSEGTGY RYYRNGGKES
151 EIQFKQNKAN GVWKQYADG SIKTEMVMVN DEPAKILTWD ESRLLSELS
201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

25
30
orf27.pep
orf27a
orf27.pep
orf27a

10 20 30
KQWYADXS IKTEMVMVNDEPAKILTWDSESG
||||| : |||||||
LSEGTGXRYRNGGKSEI QFKQNKANGVWKQWYADGN IKTEMVMVNDEPAKILTWDSESG
140 150 160 170 180 190
40 50 60 70 80
RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG YLIEP
||||| : |||||||
RLLSELSIRHHQRNGVVLEWYEDGSKKXEA VYQDDKLVRKTQWDKDG YLIEP
200 210 220 230 240

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35
40
45
1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
51 GGGCGCTTTG CCGGCGCAGA NCTATTCTGT TTATTTTAAAT CAGAACGGGA
101 AACTGACGGC GACGNTGTCT TCTGCCGCGT ATATCAGGCA ATATAGTGTG
151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
201 ATATTCGGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT
401 TGAGTGAAGG TACGGGGTNN CGCTATTACC GTACCGGCGG CAAGGAAAGC
451 GAAATCCAGT TTAACAGAAA TAAGGCAAAC GGCGTATGGA AGCAATGGTA
501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
551 CCAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

This encodes a protein having amino acid sequence <SEQ ID 702>:

50
55
1 MKKLSRIVFS TVLLGFS AAL PAQXYSVYFN QNGKLTATXS SAAYIROYSV
51 AEGIAHAQXF XYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
101 KMAGGFSK GK PDGEWVNWYP NGKKS AVMPY KNGLSEGTGX RYYRNGGKES
151 EIQFKQNKAN GVWKQYADG NIKTEMVMVN DEPAKILTWD ESRLLSELS
201 IHHHXRNGVV LEWYEDGSKK XEA VYQDDKL VRKTQWDKDG YLIEP*

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

orf27a.pep MKKLSRIVFSTVLLGFS AALPAQXYSVYFNQNGKLTATXSSAAYIROYSVAEGIAHAQXF

10 20 30 40 50 60

-402-

5	orf27-1	: :	MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF	10	20	30	40	50	60
	orf27a.pep		XYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLLWHFXGQKKMAGGFSKGPGEWVNWYP	70	80	90	100	110	120
10	orf27-1	: :	YYPSPMKKYSEPYIVASTQIKSFVPTLQNGMLLWHFNGQKKMAGGFSKGPGEWVNWYP	70	80	90	100	110	120
	orf27a.pep		NGKKSAMVPYKNGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN	130	140	150	160	170	180
15	orf27-1	: :	NGKKSAMVPYKNGLSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN	130	140	150	160	170	180
	orf27a.pep		DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKXEAQYQDDKLVRKTQWDXDG	190	200	210	220	230	240
20	orf27-1	: :	DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	190	200	210	220	230	240
	orf27a.pep		YLIEPX						
25	orf27-1		YLIEPX						

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from *N.gonorrhoeae*:

30	orf27.pep		KQWYADXS IKTEMVMVNDEPAKILTWDSE	30
	orf27ng	LSEGTGYRYYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDSE		193
35	orf27.pep	RLLSELSIRHHQRNGVVLEWYEDGSKKSEXVYQDDKLVRKTQWDKDG	YLIEP	82
	orf27ng	RLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVRKTQWDKDG	YLIEP	245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1	ATGAAGAAAT	TATCTCGGAT	TGTATTTTCA	ATCGTACTGT	TGGGTTTTTC
	51	GGCGGCTTTG	CCGCGCAGAG	CCTATCTCTG	TTATTTTAAT	CAGAACGGGA
	101	AACGTACGGC	GACGATGTCT	TCTGCCGCTT	ATATCAGGCA	ATATAGTGTG
	151	GCGGCGGGTA	TCGCACACGC	GCAGGATTTT	TATTATCCGT	CGATGAAGAA
45	201	ATATTCCGAA	CCTTATATCG	TTGCTTCAAC	GCAAATCAAA	TCTTTTGTGC
	251	CTACCCTGCA	AAACGGTATG	TTGATTTTGT	GGCATTTTAA	TGGTCAGAAA
	301	AAAATGGCGG	GGGGCTTCAG	CAAGGGTAAG	CCGGACGGGG	AATGGGTCAA
	351	CTGGTATCCG	AACGGTAAAA	AATCTGCGGT	TATGCCTTAT	AAAAATGGCT
50	401	TGAGTGAGGG	TACGGGATAC	CGTTATTACC	GTAACGGCGG	CAAGGAAAGC
	451	GAAATCCAGT	TTAAGCAAAA	TAAGGCGAAC	GGCGTATGGA	AGCAATGGTA
	501	TGCCGATGGA	AGTATCAAGA	CGGAAATGGT	TATGGTCAAC	GATGAGCCTG
	551	CCAAAATTCT	GACTTGGGAT	GAAAGCGGCC	GATTACTTTC	GGAAGTGTCT
	601	ATCCGCCACC	ATAAACGCAA	CGGGGTGGTT	TTGGAGTGGT	ATGAAGATGG
	651	TTCTAAAAAG	AGCGAGGCTG	TTTATCAGGA	TGACAAGTTG	GTCAGGAAAA
	701	CCCAATGGGA	TAAGGATGGT	TATTTAATCG	AACCTGA	

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1	MKKLSRIVFS	IVLLGFSAAL	PAQTYSVYFN	QNGKLTATMS	SAAYIRQYSV
	51	AAGIAHAQDF	YYPSPMKKYSE	PYIVASTQIK	SFVPTLQNGM	LILWHFNGQK
	101	KMAGGFSK GK	PDGEWVNWYP	NGKKSAMVPY	KNGLSEGTGY	RYYRNGGKES
	151	EIQFKQNKAN	GVWKQWYADG	SIKTEMVMVN	DEPAKILTWD	ESGRLLSELS
	201	IRHHKRNGVV	LEWYEDGSKK	SEAVYQDDKL	VRKTQWDKDG	YLIEP*

ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

	10	20	30	40	50	60
orf27-1.pep	MKKLSRIVFSTVLLGFSAALPAQTYSVYFNQNGKLTATMSSAAYIRQYSVVAGIAHAQDF					

30

35

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

40

45

50

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHX LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTTFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIIRSON ORNYVAVFAL FVLGGTHAAF

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

1  ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
5  51  TTCGCTGGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCCAC GCGGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGGGCTG GCTGCGCGGA TTGCCGCCCT TATCCCGGGT
301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTT TCTGGTACGG
10 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGCACGCA TGCGGCGTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGA CTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTTCGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTTCCAGT
15 601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACTGCCAT
651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCTTTG
701 CGCGAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTCC
20 851 TCAATCTGGG TGTGCATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CATACGGGCA ATCCGATTTA
951 TCCGCCGCCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCGTCCG TATGTTGCC GTATTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTTGTTTGA CTCGCGCTT TGGTGTATGC
25 1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSPLL WFGYTGTHE LSGFYWHAHE
30 51  MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLEFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMA HGVLAWSAV FAFAGVIET VQYRWWYKP
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGGIGVLT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
35 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

10      20      30      40      50      60
orf47.pep MKFTKHPVWAMAFRPFYSLAALYGALSPLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
45      |
orf47a    MKFTKHPVWAMAFRPFYSLAALYGALSPLLWFGYTGTHLSGFYWHAHEMIWGYAGLVV
10      20      30      40      50      60

70      80      90      100     110     120
orf47.pep IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLEFFWYGAVC
50      |
orf47a    IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLEFFWYGAVC
70      80      90      100     110     120

130     140     150     160     170
orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGLLSGLQSGLVM
55      |
orf47a    MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGLLSGLQSGLVMVSGFIGLI
130     140     150     160     170     180

60      orf47a    GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIET
190     200     210     220     230     240

```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTAA
51  TTCCTGCGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
5  151  ATGATTGGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201  CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGCG GTTCTGGTCG
251  GCTTGACTAT CTTTGGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301  TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
10  351  CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
401  ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGGCGTTC
451  CACGTCCAGC TGCACAACGG CAACCTAGGC GGA CTCTTGA GCGGATGCA
501  TCGGGGCTTG GTGATGGTGT CGGGTTTAT CGGTCTGATT GGTACGCGGA
551  TTATTTCGTT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCAGT
601  CCGAAATGGG TGGCGCAGGC TTCGTGTGG CTGCCCATGC TGACCGCCAT
15  651  TCGATGCGG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701  CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751  GTGTTGAAAG AGCCGATGCT GTGGATCTGT TTTGCCGGCT ATCTGTTTAC
801  CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAA CCCGCTTTC
851  TCAATCTGGG TGTGATCTGT ATCGGGGTCG GCGGTATCGG CGTGCTGACT
20  901  TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CACACGGGCA ATCCGATTTA
951  TCCGCGCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCGGTCGG TATGGTTGCC GTATTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
1101 GTGGAAGTAT ATTCCTTGGC TGATTCTGCC GCGTTCGGAC GGCAGGCCCG
25  1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTIFWL AARIAAFIPG
101  WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
30  151  HVQLHNGNLG GLLSGLQSG LVMVSGFIGLI GTRIIFFTS KRLNVPQIPS
201  PKWVAQASLW LPMLTAMLMA HGVMPWLSAA FAFAAGVIFT VQVYRWYKPK
251  VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
301  LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351  SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10      20      30      40      50      60
orf47a.pep  MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
|||||
orf47-1     MKFTKHPVWAMAFRPFYSLAALYGALSVLLWFGYTGTHELSGFYWHAHEMIWGYAGLVV
10      20      30      40      50      60

70      80      90      100     110     120
orf47a.pep  IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
|||||
orf47-1     IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC
70      80      90      100     110     120

130     140     150     160     170     180
orf47a.pep  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
|||||
orf47-1     MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
130     140     150     160     170     180

190     200     210     220     230     240
orf47a.pep  GTRIIFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAAGVIFT
|||||
orf47-1     GTRIIFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMAHGVLAWL SAVFAFAAGVIFT
190     200     210     220     230     240

250     260     270     280     290     300
orf47a.pep  VQVYRWYKPKVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
|||||
orf47-1     VQVYRWYKPKVLKEPMLWILFAGYLTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
250     260     270     280     290     300

310     320     330     340     350     360

```

-406-

5	orf47a.pep	LGM MARTALGHTGNPIYPPPKAVPVAFWLMM AATAVRMVAVFSSGTAYTHSIRTSSVLEA
	orf47-1	LGM MARTALGHTGNPIYPPPKAVPVAFWLMM AATAVRMVAVFSSGTAYTHSIRTSSVLEA
		310 320 330 340 350 360
		370 380
	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX
	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
10		370 380

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGVVLGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGVVLGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
	ORF47	MALPVIRSQNRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVM	172
25	ORF47ng	MALPVIRSQNRNRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGGLVMVWGFIGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

	1	<u>MKFTKHPVWA</u>	<u>MAFRPFYSLA</u>	ALYGALSVLL	WGFGYTGTHE	LSGFYWHAHE
30	51	<u>MIWGYAGLVV</u>	<u>IAFLLTAVAT</u>	WTGQPPTRGG	<u>VLVGLTAFWL</u>	<u>AARIAAFIPG</u>
	101	<u>WGAAASGILG</u>	<u>TLEFFWYGAVC</u>	<u>MALPVIRSQN</u>	<u>RRNYVAVFAI</u>	<u>FVLGGTHAAF</u>
	151	<u>HVQLHNGNLG</u>	<u>GLLSGLQSGL</u>	<u>VMVWGFIGLI</u>	<u>GMKIISFFTS</u>	<u>KRLKLPQIPS</u>
	201	<u>PKWVAHASLW</u>	<u>LPMLNAILMA</u>	<u>HRVMPWLSAA</u>	<u>FPFAAGVIFT</u>	<u>VQVYAGGITP</u>
	251	<u>IEETSCGSVA</u>	<u>GICYRLGNSS</u>	G		

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

	TM segments in ORF47ng			
40	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACTGGCG	GCACTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGGG	AACGCACGAG	CTGTCCGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTGGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
50	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCAC	GAGGGGCGGC	GTTCTGGTCG
	251	GCTTGACCGC	CTTTTGGCTG	GCTGCGCGGA	TGCGCGCCTT	TATCCCGGGT
	301	TGGGGTGGCG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CCGTTATCCG	TtcgCAAAAC	CGGCGCAACT
	401	ATGtcgCCGT	ATTGCAATA	TTTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtccAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
55	501	GTCGGGCCTG	GTTATGGTGT	CGGGCTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTTCGTT	TTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCAGT
	601	CCGAAATGGG	TGGCGCAGGC	TTGCTGTGG	CTACCCATGC	TGACCGCCAT

-407-

5 651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
 701 CGGCGGGCGT GATTTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
 751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
 851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
 901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATTCGATTTA
 951 TCCGCGCCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
 1001 CCGCCGTCCT TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
 1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
 1101 GTGGAAATAC ATTCCTGGGC TGATCCGTCC GCGTTCCGAC GGCAGGCCCG
 1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

15 1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
 51 MIWGYAGLVV IAFLLTAVAT WTGQPTRGG VLVGLTAFWL AARIAAFIPG
 101 WGAASGILG TLFFWYGAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
 151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFFTS KRLNVPQIPS
 201 PKWVAQASLW LPMLTAILMA HGVPWLSAA FAFAAGVIFT VQVYRWYKYP
 251 VLKEPMLWIL FAGYLFTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
 301 LGMMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
 20 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

		10	20	30	40	50	60
25	orf47-1.pep	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV					
	orf47ng-1	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV					
		10	20	30	40	50	60
30	orf47-1.pep	IAFLLTAVATWTGQPTRGGVLVGLTIFWLAARIAAFIPGWGASASGILGTTLFFWYGAVC					
	orf47ng-1	IAFLLTAVATWTGQPTRGGVLVGLTAFWLAARIAAFIPGWGAAASGILGTTLFFWYGAVC					
		70	80	90	100	110	120
35	orf47-1.pep	MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI					
	orf47ng-1	MALPVIRSONRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI					
		130	140	150	160	170	180
40	orf47-1.pep	GTRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAMLAHGVLAWSAVFAFAAGVIFT					
	orf47ng-1	GMRIISFFTSKRLNVPQIPSPKWVAQASLWLPMLTAILMAHGVPWLSAFAFAAGVIFT					
		190	200	210	220	230	240
45	orf47-1.pep	VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT					
	orf47ng-1	VQVYRWYKYPVLKEPMLWILFAGYLFTGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT					
		250	260	270	280	290	300
50	orf47-1.pep	LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
	orf47ng-1	LGMMARTALGHTGNSIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
		310	320	330	340	350	360
55	orf47-1.pep	LALLVYAWKYIPWLIRPRSDGRPGX					
	orf47ng-1	LALLVYAWKYIPWLIRPRSDGRPGX					
		370	380				
60	orf47-1.pep						
	orf47ng-1						
		370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
 Score = 155 bits (389), Expect = 5e-37

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

Query: 7 PVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFY-----WHAHEMIWGYAGLV 59
 P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
 5 Sbjct: 14 PIWRLAERPFPLAGSLYALLAIPLWVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
 V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF
 10 Sbjct: 72 VAGFLLTAVQTWTGQTAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPLDLFLVALVW 130

Query: 120 CMALPVIRSQNRNRYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
 MA + + +RNY V + ++ G +V+ + L
 15 Sbjct: 131 MMAQMLWAVRQKRNYPIVVVLSLMLGADVLILTGLLQGNDAQRQGVLAGLWLVAAIMMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAAFAFA 234
 IG R+I FFT + L P W+ A L + A+L A GV P L F A
 20 Sbjct: 191 IGGRVIFFFTQRLGKVDVAVKPVWLDVALLVGTGVIALLHAFGVAMRPQPLGLLFFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLEPMLWILFAGYLFGLGLIAGVASYF-KPAFXXXXXXXXXXX 293
 GV +++ RW+ K + K +LW L L+ + + +F A
 25 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLVVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPVAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353
 M+AR LGHTG + P + AF L F S +
 30 Sbjct: 310 GSMSGLILAMIAVTLGHTGRPLQPLAGIIG-AFVL---FNLGTAARVFLSVAVPVGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
 ++V + LA +Y W+Y P L+ R DG PG
 35 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

35 1 ..ATGCCGCTCTG AAGGTTTCAGA CGGCmTCGGT GyCGGGGAAY CAGAAGyGGT
 51 AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
 101 CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
 151 GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
 201 TGTkGCTTTC GTGATAGGsA GGTtTGyTGG knksAsyTTG TAyrATwkkG
 40 251 CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTtTy
 301 TkTtTyCACc GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
 351 CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
 401 GACAAGCCGA GAGAAGAAAC GGCCTGGAAG CTGCCGTTTC CCTGATGTTT
 451 TGTTTGGGTT TCTTTGTAGT TGTGTTTAT CTCTTCAGTA ACTTTTTTAG
 45 501 TAGAAGAATT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCCGCT
 551 ATTCTCCAGC CGCCGAAATC ..

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

50 1 ..MPSEGSdGXG XGEXEVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX
 51 DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXSRGF
 101 XXHRMNLmFN VSVGDARADI GFEFIVEFEI VNGGQAERRN GVEAAVSLMF
 151 CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

	orf67.pep				MPSEGS	DXGXGEXEXVAHAQXDFVGF	EAG	30
	orf67ng	TNFEI	AVLSG	MTVRV	FYCAR	PAPVNG	GRGRLK	MPSEGS
		90	100	110	120	130	140	146
5	orf67.pep	V	FQAS	PVV	TVSG	VXXQL	GXD	VETD
	orf67ng	V	FQAS	PVV	VAVAG	VQGQ	AGR	DVYA
10	orf67.pep	X	WXXX	SRG	FXXH	RMLM	FNV	SVGD
		:	:	:	:	:	:	:
	orf67ng	T	RVGK	STC	YFFS	RIDA	VSDV	SVGD
15	orf67.pep	C	LGEF	FVV	-----	V	VYLF	SNFF
					:	:	:	:
	orf67ng	R	LLV	FYV	KLV	AAK	SFI	LSF

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

	1	MPSET	VGSIV	NVG	VDES	VGF	SPP	FPSI	QHF	YRF	HRI	HRIR	LFR	PPG	PMQL
20	51	NRH	SHG	SGNL	GRG	VWAT	VLS	DKF	PCG	QVRI	PAC	AGM	TNFE	IAV	LSG
	101	VFY	CAR	PAPV	NGG	R	LKMP	SE	GSD	GIG	IGES	EAV	AHA	Q	RGF
	151	SPV	V	VAVAG	Q	QAG	R	DVYA	HAR	HRAE	AQA	AAV	A	FLIG	V
	201	NCC	V	SITRVG	G	KST	C	YFFSR	ID	A	VSD	V	SVG	DART	D
	251	QAE	R	NGVEC	AV	FLM	FRL	LLV	FYV	KLV	AAK	FIL	S	FQ	LFY
25	301	PVT	G	I	R	GDA	PAA	E	V	V	A	D	R	H	P
	351	I	V	G	N	A	F	G	G	V	G	*			

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATG	TTT	GCTT	TTT	AGA	AGC	CTT	TTT	TGTC	GA	A	TAC	G	G
	51	TTT	TTT	TGTA	TTG	GC	T	CAT	CT	GCG	G	T	T	CGG	CG
35	101	TG	AC	CT	T	G	G	T	T	G	T	T	CGG	GT	A
	151	CAT	ATT	TAT	GT	TGC	AG	T	CGG	TAT	G	C	T	CGG	C
	201	CAT	G	T	T	C	G	C	C	G	A	C	G	A	A
	251	CT	ATT	T	G	C	G	S	G	CAT	C	A	T	G	A
	301	TTC	G	A	C	A	A	A	T	AC	G	G	T	A	A
	351	TTT	G	A	A	A	C	G	T	T	T	T	G	T	T
40	401	ACT	T	G	C	G	T	T	T	TAT	C	A	T	A	T

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFA	F	L	E	A	F	F	V	EY	G	A	A	V	F	F
	51	HIM	F	A	V	G	M	L	G	V	L	V	G	D	G	I
	101	F	D	K	Y	G	N	W	V	L	F	A	R	L	P	L

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATG	TTT	GCTT	TTT	AGA	AGC	CTT	TTT	TGTC	GA	A	TAC	G	G
	51	TTT	TTT	TGTA	TTG	GC	T	CAT	CT	GCG	G	T	T	CGG	CG
	101	TG	AC	CT	T	G	G	T	T	G	T	T	CGG	GT	A
	151	CAT	ATT	TAT	GT	TGC	AG	T	CGG	TAT	G	C	T	CGG	C
	201	CAT	G	T	T	C	G	C	C	G	A	C	G	A	A
	251	CT	ATT	T	G	C	G	S	G	CAT	C	A	T	G	A
	301	TTC	G	A	C	A	A	A	T	AC	G	G	T	A	A
	351	TTT	G	A	A	A	C	G	T	T	T	T	G	T	T
	401	ACT	T	G	C	G	T	T	T	TAT	C	A	T	A	T
55	451	ATT	T	G	G	A	T	T	T	AT	C	G	G	G	A

```

501  GCGCAAAATG CACAGCCTGC AATCGGGTAT TTTGTTATC TTGGGTATAG
551  GTGCGACCGT TGTGCTTGG ATTTGGTGA AAAACGCCA ACGTATCCAG
601  TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
651  GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

```

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```

1  MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51  HIMFAVGMLG VLVDGIMFA AGRIWGQKIL RFKPIARIMT PKRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVEVTAGISR KVSYLRFIIM DGLAALISVP
151 IWIYLGEYGA HNIDWLMAKM HSLQSGIFVI LGIGATVVAW IWWKKRQRIQ
201 FYRSKLKEKR AQRKAATAKA KAAQSKQ*

```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

15  Orf78: 4   FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
      FL  FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+      N H+M V M+GV
      DedA: 20 FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHMLLVSMIGV 79

20  Orf78: 62   LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
      L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR
      DedA: 80 LAGDSCMYWLGRIYGTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

      Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
      +++ +GI+R+VSY+RF+++D AA
25  DedA: 140 IYMVSGITRRVSVYRVFLIDFCAA 163

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

```

30  orf78.pep      10      20      30      40      50      60
      MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
      orf78a      10      20      30      40      50      60
      MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG

35  orf78.pep      70      80      90      100     110     120
      VLVDGIMFAAGRIWGQXXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRT
      orf78a      70      80      90      100     110     120
      VLVDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT

40  orf78.pep      130     140
      AVFVTAGISRKVSYLRFIIMDGLAA
      orf78a      130     140
      AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA

45  orf78a      130     140     150     160     170     180
      AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIA

```

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

```

1  ATGTTTGCCC TTTTGAAGC CTTTTTTGTC GAATACGGCT ATGCGGCCGT
51  GTTTTTTCGT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101 TGACCTTGGT AACAGGCGGC GTGATTTCCG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201 CATGTTCCGC GCCGGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
301 TTCGACAAAT ACGGCAACTG GGTGTTATTT GTCGCTCGTT TCCTGCCCGG
351 TTTGCGGACT GCCGTTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT

```

-411-

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTTCATCGCA TTGGGCGTGC
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
 51 HIMFAVGMLG VLVGDGIMEA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFILM DGLAALISVP
 151 VWIYLGEYGA HNIDWIMAKM HSLQSGIFIA LGVLAALAW FWRKRRHYQ
 201 LYRAQLSEKR AKRKA EAKAAK KAAQKQ*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
orf78a.pep		MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
orf78-1		MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG					
		10	20	30	40	50	60
orf78a.pep		VLVGDGIMFAAGRIWGQKILKFKPIARIMT PKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
orf78-1		VLVGDGIMFAAGRIWGQKILRFKPIARIMT PKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
orf78a.pep		AVFVTAGISRKVSYLRFILMDGLAALISVPVWVWYIYLGEYGAHNIDWIMAKMHSLQSGIFIA					
orf78-1		AVFVTAGISRKVSYLRFIIMDGLAALISVPIWVWYIYLGEYGAHNIDWIMAKMHSLQSGIFVI					
		130	140	150	160	170	180
orf78a.pep		LGVLAALAWFWRKRRHYQLYRAQLSEKRKAERKA EAKAAKAAQKQX					
orf78-1		LGIGATVVAWIWKKRQRIQFYRSKLKEKRAQRKA EAKAAKAAQSKQX					
		190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

orf78.pep	XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
orf78.pep	IIMDGLAA	145
orf78ng	LIMDGLAALISVPVWVWYIYLGEYGAHNIDWIMAKMHSLQSGIFIALGVLAALAWFWRKRR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

1 ..YPVLFVARFL PGLRTAVFT AGISRKVSYL RFLIMDGLAA LISVPVWYIYL
 51 GEYGAHNIDW LMAKMHSLQS GIFIALGVLA AALAWFWRK RRHYQLYRAQ
 101 LSEKRAKRKA EKAAKAAQK QQ*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

1 atgtttgccc tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT
 51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
 101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTGGCGG GCGACGGCGT
 201 GATGTTTGCC GCCGACGCA TCTGGGGGCA GAAAATCCTC AAGTTCAAAC
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGCAACTG GGTCTGTTT GTCGCCCGTT TCCTGCCGGG

-412-

5 351 TTTGCGGACT GCCGTTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
 401 ATCTGCGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCC
 451 GTTTGGATT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT
 501 GGCAGAAATG CACAGCCTGC AATCGGGCAT CTTTCATCGCA TTGGGCGTGC
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10 1 MFALLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTlVTGG VISGMGYTNP
 51 HIMFAVGMLG VLAGDGMFA AGRIWQKIL KFKPIARIMT PKRYAQVQEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRELIM DGLAALISVP
 151 VWIYLGEYGA HNIDWLMAM HSLQSGIFIA LGVLAALAW FWRKRHHYQ
 201 LYRAQLSEKR AKRKAKEAAK KAAQKQ*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15 orf78-1.pep 10 20 30 40 50 60
 orf78ng-1 MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNP
 20 10 20 30 40 50 60
 orf78-1.pep 70 80 90 100 110 120
 orf78ng-1 VLAGDGMFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT
 25 70 80 90 100 110 120
 orf78-1.pep 130 140 150 160 170 180
 orf78ng-1 AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMHSLQSGIFIA
 30 130 140 150 160 170 180
 orf78-1.pep 190 200 210 220
 orf78ng-1 LGVLAALAWFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX
 35 190 200 210 220

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

40 sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA
 protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
 >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
 Score = 223 bits (563), Expect = 7e-58
 Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)
 45 Query: 5 LEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGM--GYTNP
 HIMFAVGMLGVL 62
 L FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GVL
 Sbjet: 21 LIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGGVIAGLYPENVN
 SHMLLVSMIGVL 80
 50 Query: 63 AGDGMFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRTAV 122
 AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +
 Sbjet: 81 AGDSYWLGRIGYKILRFPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140
 55 Query: 123 FVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAMHSLQSGIFIALG 182
 ++ +GI+R+VSY+RF+++D AA+ISVP+WIIYLGE GA N+DWL ++ Q I+I +G
 Sbjet: 141 YMVSGITRRVSYVRFVLIDFCAAIISVPIWIYLGEYGAHNIDWLMAMHSLQSGIFIALG 200
 Query: 183 VL 184
 L
 Sbjet: 201 YL 202
 60

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

10

```

      1  ATGAAAAAAT  TATTGCGCGC  CGTGATGATG  GCAGGTTTGG  CAGGCGCGGT
     51  TTCCGCCGCC  GGAGTCCACG  TTGAGGACGG  CTGGGCGCGC  ACCACCGTCG
    101  AAGGTATGAA  AATAGCGCGC  GCGTTCATGA  AAATCCACAA  CGACGAAGCC
    151  AAACAAGACT  TTTTGCTCGG  CGGAAGCAGC  CCCGTCGCG  ACCGCGTCGA
    201  AGTGCATACC  CACATCAACG  ACAACGCGCT  GATGCGGATG  CGCGAAGTCG
    251  AAGGCGGCGT  GCCTTTGGAA  GCGAAATCCG  TTACCGAACT  CAAACCGGCG
    301  AGCTATCATG  TGATGTTTAT  GGGTTTGAAG  AAACAATTAA  AAGAGGGCGA
    351  TAAATTTCCC  GTTACCCTGA  AATTTAAAAA  GCCCAAGACG  CAAACCGTCC
    401  AACTGGAAGT  CAAAATCGCG  CCGATGCCGG  CAATGAACCA  C...
```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHNDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KOLKEGDKIP VTLKFKNKA QTQLEVKIA PMPAMNH..

```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

20 1 ATGAAAAAAT TATTGCGGCG CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATAGCGGCG GCGTTCATGA AAATCCACAA CGACGAAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCCGTTGCCG ACCCGTGCGA
201 AGTGCATACC CATTCAACAG ACAACGCGGT CGCGAAGTCG
25 251 AAGGCGGCGT GCCTTTGGAA GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TGATGTTTAT GGGTTTGAAA AAACAATTAA AAGAGGGCGA
351 TAAAAATCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
401 AACTGGAAGT CAAAATCGCG CCGATGCCGG CAATGAACCA CGGTATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

- 30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHND
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKIA PMPAMNHGHH
151 HGEAHOH*

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N. meningitidis*:

40		10 20 30 40 50 60
	orf79.pep	MKKLLAAVMMAGLAGAVSAAGVHVEDGWARTTVEGMKIGAFMKIHNDEAKQDFLLGGSS : : : :
	orf79a	<u>MKKLLAAVMMAGLAGAVSAAGIHVEDGWARTTVEGKMKGAFMKIHNDEAKQDFLLGGSS</u>
45		10 20 30 40 50 60
	orf79.pep	PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVFMFGLKKQLKEGDKIP : : : :
	orf79a	PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVFMFGXKKQLKXGDKIP
50		70 80 90 100 110 120

-414-

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAPMPAMNH
              |||||
5 orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHHGEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGTCTCG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
201 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA
151 351 CAAGATCCC GTTACCCTGA AATTAAAAA CGCCAAAGCA CAAACCGTCC
401 AACTGGAAGT CAAACCCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TTVEGMKMG GAFMKIHNDEA
20 51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25 orf79a.pep    MKXLLAAVMMAGLAGAVSAAGIHHVEDGWAR TTVEGMKMG GAFMKIHNDEAKQDFLLGGSS
    || |||||
orf79-1         MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIG GAFMKIHNDEAKQDFLLGGSS
              10      20      30      40      50      60
30 orf79a.pep    PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKQLKXGDKIP
    |||||
orf79-1         PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
              70      80      90      100     110     120
35 orf79a.pep    VTLKFKNAKAQTVQLEVKTAPMSAMDHGHGHHGEAHQH
    |||||
orf79-1         VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGHHGEAHQH
              130     140     150
40

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45 orf79.pep    FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
    |||||
orf79ng         INDNGVMRMREVKGVPLEAKSVTELKPGS 30
50 orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
    |||||
orf79ng         YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHGHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

55 1  ..INDNGVMRM EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
51 TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

1  ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51  TTccgccgCc GGagTccAtG TCGaggACGG CTGGGCGCGc accaCTGtcg
101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
151 atacaaGACT ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTGCA
5  201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACCTGA AAGAGGGCGA
351 CAAGATTCCC GTTACCCTGA AATTTAAAAA CGCCAAAGCG CAAACCGTCC
10 401 AACTGGAAGT CAAAACCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

1  MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMGG AFMKIHNDDEA
51  IQDFVLGGSM PVADRVVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
15  101 SYHVMFMGLK KQLKEGDKIP VTLKFNAKA QTVQLEVKTA PMSAMNHGHH
151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

                10      20      30      40      50      60
orf79-1.pep  MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDDEAKQDFLLGGSS
20  orf79ng-1  MKKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKMGGA FMKIHNDDEAIQDFVLGGSM
                10      20      30      40      50      60

                70      80      90      100     110     120
orf79-1.pep  PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
25  orf79ng-1  PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
                70      80      90      100     110     120

                130     140     150
30  orf79-1.pep  VTLKFNAKAQTVQLEVKIAMPAMNHGHHHGEAHQH*
                130     140     150
orf79ng-1  VTLKFNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH*
                130     140     150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35  gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
    Score = 63.6 bits (152), Expect = 6e-10
    Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

40  Query: 24  VEDGWAR TTVEGMKMGGA FMKIHNDDEAIQDFVLGGSM PVADRVVHTHINDNGVMRMREV 83
    V+  W      G      M I N+  D+++G  +A RVE+H  + +N V +M
    Sbjct: 27  VKHPVME PPPGPNTTMMGMII VNEGDEPDYLIGAKTDIAQRVELHKTVIENDVAKMVPQ 86

    Query: 84  KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFNAKAQTVQLEV 137
    + + + K  E K  YHVM +GLKK++KEGDK+ V L F+ +  TV+  V
45  Sbjct: 87  ER-IEIPPKGKVEFKHHGYHVMII GLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

5      1  ATGACGGTAA CTGCGGCCGA AGGCGGCAAA GCTGCCAAGG CGTTAAAAAA
      51  ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
     101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
     151  CCGAAGCAAT GGCGGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
     201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGA TTGTTTGCCG
     251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
    10  301  CGGATTCCGG TTGTGAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
     351  ATAcgTGCTG TCCGACAGCA GCCGTTTCGT TAAAACGCCG GTACTCGTGC
     401  CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTCGTGTC AGGGCAGGTG
     451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAs GACGGCGATT ATCTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGTTACTAT ATTATGGTAA
    15  551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AsCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
     701  AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20      1  MTVTAAEGGK AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL
     51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLlG
    101  RIPVVKSIYS SVKKVSEYVL SDSSRSFKTP VLVFPFPQGI WTIAFVSGQV
    151  SNAVKAALPX DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEXLK
    201  YVISLGMVIP DDLPVKTLAX PMPSEKADLP EQQ*

```

25 Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

      1  ATGACGGAAC nTGCGGCCGA AGGCGGCAAA GCTGCCAArG CGTTAAAAAA
     51  ATATCTGATT ACGGGCATTT TGGTCTGGCT GCCGATTGCG GTAACGGTTT
    101  GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
    151  CCGAAGCAAT GGCGGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
    201  GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAAACCGA TTGTTTGCCG
    251  CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCTGTTGGGG
    301  CGGATTCCGG TTGTGAAATC CATCTATTCG AGTGTGAAAA AAGTATCCGA
    351  ATCGCTGCTG TCCGACAGCA GCCGTTTCGT TAAAACGCCG GTACTCGTGC
    401  CGTTTCCCCA GCCCGGTATT TGGACGATTG CTTCGTGTC AGGGCAGGTG
    35  451  TCGAATGCGG TTAAGGCCGC ATTGCCGAAG GACGGCGATT ATCTTCCGT
     501  GTATGTTCCG ACCACGCCGA ATCCGACCGG CGTTACTAT ATTATGGTAA
     551  AGAAAAGCGA TGTGCGCGAA CTCGATATGA GCGTGGACGA AGCATTGAAA
     601  TATGTGATTT CGCTGGGTAT GGTCAATCCCT GACGACCTGC CCGTCAAAAC
     651  ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACAACAAT
    40  701  AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

      1  MTEXAAEGGK AAKALKKYLI TGILVWLPiA VTVWVVSyIV SASDQLVNLL
     51  PKQWRPQYVL GFNIPGLGVI VAIaVLFVTG LFAANVLGRQ ILAAWDSLlG
    101  RIPVVKSIYS SVKKVSESL SDSSRSFKTP VLVFPFPQGI WTIAFVSGQV
    45  151  SNAVKAALPK DGDYLSVYVP TTPNPTGGYY IMVKKSDVRE LDMSVDEALK
    201  YVISLGMVIP DDLPVKTLAg PMPSEKADLP EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N. meningitidis*:

```

50      orf98.pep      10      20      30      40      50      60
      orf98a      MTVTAAEGGKA AKALKKYLI TGILVWLPiAVTVWVVSyIVSASDQLVNLLPKQWRPQYVL
      II  |||||||||||||||||||||||||||||||||||||||||||||||||||
55      orf98a      MTEPAAEGGKA AKALKKYLI TGILVWLPiAVTVWVVSyIVSASDQLVNLLPKQWRPQYVL
      10      20      30      40      50      60

```

-417-

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
5	orf98a	GFNIPGLGVIVAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVXSLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
10	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
	orf98a	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
15	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAKPMPSEKADLPEQQX					
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTAATA	TCCCGGGGCT
25	201	GGCGGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCCG
	251	CAAACGTATT	GGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTTGTGTTGGG
	301	CGGATTCGG	TTGTGAAGTC	CATCTATTG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTG	TAAACACCA	GTAATCGTGC
	401	CGTTTCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG
30	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKVSXSLL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV
40	151	SNAVKAALPK	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDLVPVKTL	LAGPMPSEK	ADLPEQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKA	AKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	PKQWRPQYVL
45	orf98-1	MTEPAAEGGKA	AKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	PKQWRPQYVL
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf98a.pep	GFNIPGLGVIVAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVXSLL					
	orf98-1	GFNIPGLGVIVAIAVLFTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSESL					
		70	80	90	100	110	120
55	orf98a.pep	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
	orf98-1	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
60	orf98a.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX					
	orf98-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLVPVKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	
65							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
5	orf98.pep	MTVTAAEGGKAAKALKKYLI	TGILVWLPIAVTVWVVS	YIVSASDQLVNLLPKQWRPQYVL				60
	orf98ng	MTEPAAEGGKAAKALKKYLI	TGILVWLPIAVTVWVVS	YIVSASDQLVNLLPKQWRPQYVL				60
	orf98.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVSEYVL					120
10	orf98ng	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLXRIIPVVKSIYSSVKKVSESL					120
	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQV	SNVKAALPXDGDYLSVYVPTTNPNTGGYY					180
15	orf98ng	SDSSRSFKTPVLVFPFQSGIWTIAFVSGQV	SNVKAALPDGDYLSVYVPTTNPNTGGYY					180
	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIP	DDLVPVKTLAXPMPSEKADLPEQQ					233
	orf98ng	IMVKKSDVRELDMSVDEALKYVISLGMVIP	DDLVPVKTLAGPMPPEKAELPEQQ					233

20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX	
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV	
25	151	SNVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK	
	201	YVISLGMVIP	DDLVPVKTLAG	PMPPEKAELP	EQQ*		

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA	
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT	
30	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG	
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTAATA	TCCCCGGGCT	
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTGCGG	
	251	CAAACGTGTT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTgggg	
35	301	cggattCCGG	TTGTCAAATC	CATCTATTCT	AGTGTGAAAA	AAGTATCCGA	
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCGTT	TAAAACGCCG	GTAACCTGTC	
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCGTGTC	CGGTCAGGTG	
	451	TCGAATGCCG	TTAAGGCCGC	ATTGCCGCG	GATGGCGATT	ATCTTCCGT	
	501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTACTAT	ATTATGGTAA	
40	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA	
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAC	
	651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCC	GAACAACAAT	
	701	AA					

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	
45	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG	
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSGQV	
	151	SNVKAALPQ	DGDYLSVYVP	TPNPNTGGYY	IMVKKSDVRE	LDMSVDEALK	
	201	YVISLGMVIP	DDLVPVKTLAG	PMPPEKAELP	EQQ*		

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

		10	20	30	40	50	60	
50	orf98-1.pep	MTEXAAEGGKAAKALKKYLI	TGILVWLPIAVTVWVVS	YIVSASDQLVNLLPKQWRPQYVL				
	orf98ng-1	MTEPAAEGGKAAKALKKYLI	TGILVWLPIAVTVWVVS	YIVSASDQLVNLLPKQWRPQYVL				
		10	20	30	40	50	60	
55	orf98-1.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVSESL					
		10	20	30	40	50	60	
		70	80	90	100	110	120	
		GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKKVSESL					

-419-

	orf98ng-1	GFNIPGLGVIVAI	AVLFV	TGLFA	ANVLGR	QILAAW	DSLLGR	IPVVK	SIYSS	VKKVSE	SL
		70	80	90	100	110	120				
5	orf98-1.pep	SDSSRSFKTPVL	VPFPQ	PGIWT	IAFVSG	QVSN	AVKAA	LPKDG	DYLSV	YVPTT	PNPTGG
	orf98ng-1	SDSSRSFKTPVL	VPFPQ	SGIWT	IAFVSG	QVSN	AVKAA	LPQDG	DYLSV	YVPTT	PNPTGG
10	orf98-1.pep	IMVKKSDVREL	DMSVDE	ALKYIS	LGMP	IPDDLP	VKTLA	GPMP	SEKAD	LPEQQX	
	orf98ng-1	IMVKKSDVREL	DMSVDE	ALKYIS	LGMP	IPDDLP	VKTLA	GPMP	PEKAE	LPEQQX	

- 15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

	1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GgGgTACTCA
25	201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCnCGTAA	AGGCCkCAAG
	251	ssCGsGCTTG	CCTTGAACAA	GGCGGTTTG	GCGTATTTTG	AAGGGCGTTT
	301	TGAAAAGGCG	GAAGTAGAAG	CCTCACGCGT	GTGGTCAAC	AAAGtAGGCC
	351	GaGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGCGCAC	GCCGCCGGAC
	401	AGATGGAAAA	CATCGAssTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
30	451	TTCCCGGAAA	AACAGCAGCT	TTCCCGTTAT	CTTTGTGTGG	CGGAATCGGC
	501	GTTGAACCGG	CGCGATTACG	AAGCGCGCGA	AGCCAATCTT	CATGCGGCGG
	551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTCGTTAC
	601	GCTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAAA	CCGAAAAACT
	651	TTCCAAGGCG	GGCGCGTTGG	GCAAATCGGA	AATGGAACGG	TATCAAAATT
35	701	GGGCATATCC	GTGCGCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
	751	AACCTGCCTG	AAGCGGATTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
	801	TATCGGTTGC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
	851	AAATGGGTCA	AACAGCATTA	TCCGCAsAAC	CGCCGCCCCG	AGCTTTTGGA
	901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
40	951	CCATCGATTT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCGA	TAACGCGCTT
	1001	CTGCTGATGT	ATCTCGGTGC	GCTCGCCTTC	GGCCGCAAAC	TTTGGGGCAA
	1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTGC	ATTAAAGCCG	AGTATTTCCG
	1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
	1151	GCGGAGGCGC	AC...			

- 45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

	1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	KMQRFSGSARK	GKXXLALNK	AGLAYFEGRE
	101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGQMenIXXR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
50	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	XNRRPELLEA
	301	FVESVRFLGE	REQQKAIDFA	DAWLKEQPDN	ALLMYLGRIL	AFGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCGTACTCAA

-420-

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA
 5 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
 10 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAATTTC
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
 701 CATACCGCCG CCACTGGCGG GATGCTGCCG ATGCCGCGCG TTTGAAAACC
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACGGGGAAT TGAGCGTATC
 801 GGTTCGCGAA AAGTACGAAC GTTGGGGACT GTATGCCGAT GCGGTCAAAT
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
 15 901 TTTGTCGAAA GCGTCGCGTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
 951 CGATTTTGCC GATGCTTGGC TGAAGAACA GCCCGATAAC GCGCTTCTGC
 1001 TGATGTATCT CCGTCGGCTC GCCTACGGCC GCAAACTTTG GGGCAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
 1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACCG CAGAAGGCGG
 20 1151 AGGCGCAGCG CAACTTGTTT TTGAAGCCG TCTCCGATGA CGAACGTCAC
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAVGLALA SGIYTG DVYI VLGQTMLRIN LHAFVLGSLI
 25 51 AVVVWYFLFK FIIGVNLNPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENTIELR DRYLAELIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNANLT RLVLRLQRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNNRRPELLEA
 301 FVESVRFLGE REQKAIDFA DAWLKEQPDN ALLMYLGRIL AYGRKLWGKA
 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEQRNLV LEAVSDDERH
 401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

		10	20	30	40	50	60
orfl00.pep		MKTVVWIVVLFAAAVGLALASGIYTG	DVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK				
orfl00a		MKTVVWIVVLFAAAXGLALASGIXTG	DVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK				
40		10	20	30	40	50	60
		70	80	90	100	110	120
orfl00.pep		FIIGVNLNPEKMQRFGSARKGKXXLALNKAGLAYFEGRF	EKAELEASRVLVNKGVRDNR				
orfl00a		FIIGVNLNPEKMQRFGSARKGRKAALALNKAGLAYFEGRF	EKAELEASRVLVNKEAGDNR				
		70	80	90	100	110	120
		130	140	150	160	170	180
orfl00.pep		TLALMLXAHAAGQMENTIXRDRYLAEIAKLPEKQQLSRYLLA	ESALNRRDYEAEEANLH				
orfl00a		TLALMLGAHAAGQMENTIELRDRYLAEIAKLPEKQQLSRYLLA	ESALNRRDYEAEEANLH				
		130	140	150	160	170	180
		190	200	210	220	230	240
orfl00.pep		AAAKMNANLTRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
orfl00a		AAAKMNANLTRLVRLQRLYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
		190	200	210	220	230	240
		250	260	270	280	290	300
orfl00.pep		DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYH	PXNRRPELLEA				
orfl00a		DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYH	PHNRRPELLEA				
		250	260	270	280	290	300

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKEQPDNALLMYLGRILAFGRKLWGKAKGYLEASIAL					
5	orf100a	FVESVRFLGERDQQKAIDFADAWLKEQPDNALLXYLGRILAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPSISARLVLTQVFDEIGEPOKAEAH					
10	orf100a	KPSISARLVLAQVFDETGEPQKAEQARNLVLASVAEENRPSAETHX					
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

15	1	ATGAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CNNTCGGGCT
	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCCG	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	GCGTACTCAA
	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAAGGCGG	AACTTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG
	451	CCGGAAAAGC	AGCAGCTTTC	CCGTTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTCGACAGGG	GCGACGCGTT	GCAGGTTCTG	GCAAAAACCG	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAAACC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAAT	TTTGAAGCN
	901	TTTGTGAAA	GCGTGCCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGGTGCGCTC	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGCTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVVYFLFK	FIIGVLNXPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAALKT
45	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYH	HNRRPELLEA
	301	FVESVRFLGE	RDQQKAIDFA	DAWLKEQPDN	ALLLXYLGR	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLV	AKVFDETGEP	QKAEQARNLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50		10	20	30	40	50	60
	orf100a.pep	MKTVVWIVVLFAAAXGLALASGIXTGDVYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
	orf100-1	MKTVVWIVVLFAAAGVLALASGIYTGVDYIVLGQTMLRINLHAFVLGSLIAVVVWYFLFK					
55		10	20	30	40	50	60
	orf100a.pep	FIIGVLNXPEKMQRFGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLGNKEAGDNR					
	orf100-1	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFKAELEASRVLVNKEAGDNR					
60		70	80	90	100	110	120
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
65		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					
	orf100-1	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLAESALNRRDYEAEEANLH					

-422-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf100a.pep	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKXSKAGAXGKSEMERYQNWAYRRQLX					
	orf100-1	AAAKMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
		190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLF	AAAVGLALASGIY	TG	DVYIVLG	QTMLRINLHAFV	LGSLIAVVVWY	FLFK	60		
	orf100ng	MKTVVWIVVLF	AAAVGLALASGIY	TG	DVYIVLG	QTMLRINLHAFV	LGSLIAVVVWY	FLFK	60		
35	orf100.pep	FIIGVLN	IP	EKMQRFGS	SARKG	XXXXLALN	KAGLAYFE	GRFEKAELEAS	RVLVNKVGR	DNR	120
	orf100ng	FIIGVLN	IP	ENMRRSGS	SARKGR	KAAALALN	KAGLAYFE	GRFEKAELEAS	RVLGNKEAG	DNR	120
40	orf100.pep	TLALMLX	AHAAGQ	MENIXXR	DRYLAE	IAKLPEK	QQLSRY	LLLAES	ALNRRDYE	AAEANLH	180
	orf100ng	TLALMLG	AHAAGQ	MENIELR	DRYLAE	IAKLPEK	QQLSRY	LLLAES	ALNRRDYE	AAEANLH	180
45	orf100.pep	AAAKMN	ANLTRL	VRLXIR	YAFDRG	DALQVL	AKTEKLS	KAGALG	KSEMERYQ	NWAYRRQLA	240
	orf100ng	AAAKMN	ANLTRL	VRLQLR	YAFDRG	DALQVL	AKTEKLS	KAGALG	KSEMERYQ	NWAYRRQMA	240
50	orf100.pep	DAADAA	ALKTCL	KRIPD	SLKNGEL	SVSVAE	KYERLGL	YADAVK	WVKQHY	PXNRRPELLEA	300
	orf100ng	DAADAA	ALKTCL	KRIPD	SLKNGEL	SVSVAE	KYERLGL	YADAVK	WVKQHY	PHNRRPELLEA	300
55	orf100.pep	FVESVR	FLGERE	QQKAID	FADAWL	KEQPDN	ALLMYL	GRLAYG	RKLWGK	AKGYLEASIAL	360
	orf100ng	FVESVR	FLGERE	QQKAID	FADSWL	KEQPDN	ALLMYL	GRLAYG	RKLWGK	AKGYLEASIAL	360
60	orf100.pep	KPSISAR	LVLTKV	FDEIGE	PQKAEAH	386					
	orf100ng	KPSIPAR	LVLAKV	FDETAO	SOKAEA	ORNLVL	ASVAGEN	RPSAETR	405		

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT	
60	51	GGCGCTGGCT	TCGGGCATT	ACACCGCGCA	CGTGATATC	GTA	CTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TT	CGCTGATT
	151	GCCGTCGTGG	TGTGGTATT	CCTGTTTAAA	TTCATCATCG	GCGT	ACTCAA
	201	TATCCCGGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCC	GCAAGG
	251	CCGCGCTTGC	CTGAATAAG	GCGGGTTTGG	CGTATTTCTGA	AGGG	CGTTT
65	301	GAAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAG	GCCCGG
	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGG	ACAGA
	401	TGGAATAAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAA	ACTG

-423-

5 451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
 501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
 601 TTCGATCGGG GCGATGCGTT GCAGGTTCTG GCAAAAaccG AAAAATTTC
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
 701 CATACCGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
 751 TGCCCTGAAGC GGATTCCTCCGA CAGCCTCAAA AACGGGGAAT TGagcGTATC
 801 GGTGTCGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
 10 901 TTTGTCGAAA GCGTGCCTTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
 951 CGATTTTGCC GATTCTTGCC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
 1001 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACTTTG GGGTAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG
 1101 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
 15 1151 AAGCACAGCG CAACTTGGTT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT
 1201 TCCGCCGAAA CCCGTTGA

This encodes a protein having amino acid sequence <SEQ ID 756>:

20 1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLQQTMLRIN LHAFLVLSLI
 51 AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
 101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEANLH AAKMNNANLT RLVRLQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWVKQHYP HNRRLPELLEA
 301 FVESVRFLGE REQKKAIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA
 25 351 KGYLEASIAL KPSIPARLVL AKVFDETAQS QKAEAQRLV LASVAGENRP
 401 SAETR*

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
30	orf100-1.pep	MKTVVWIVVLFAAAVGLALASGIYTGdVYI	VLQQTMLRINLHAFLVLSLI	AVVVWYFLFK			
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGdVYI	VLQQTMLRINLHAFLVLSLI	AVVVWYFLFK			
		10	20	30	40	50	60
35	orf100-1.pep	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRF	EKAELEASRV LGNKEAGDNR				
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRF	EKAELEASRV LGNKEAGDNR				
		70	80	90	100	110	120
40	orf100-1.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLL	LLAESALNRRDYEAAEANLH				
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLL	LLAESALNRRDYEAAEANLH				
		130	140	150	160	170	180
45	orf100-1.pep	AAAKMNNANLRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERY	QNWAYRRQMA				
	orf100ng	AAAKMNNANLRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERY	QNWAYRRQMA				
		190	200	210	220	230	240
50	orf100-1.pep	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYP	HNRRLPELLEA				
	orf100ng	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWVKQHYP	HNRRLPELLEA				
		250	260	270	280	290	300
55	orf100-1.pep	FVESVRFLGEREQKKAIDFADAWLKEQPDNALLMYLGR LAYGRKLWGKAKGYLEASIAL					
	orf100ng	FVESVRFLGEREQKKAIDFADAWLKEQPDNALLMYLGR LAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
60	orf100-1.pep	KPSISARLV LAKVFDEIGEPQKAEAQRLV LASVAGENRPSAETR					
	orf100n	KPSIPARLV LAKVFDETAQSQKAEAQRLV LASVAGENRPSAETR					
		370	380	390	400		

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
      51  GTTTCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
      101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
      151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTG GCGTTCGGCG CGGTCGTGTT
      201 CCGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
      251 ACGTCAAACGT GTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
      15  301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
      351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401 TGTATSTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRIEVD MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCIG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYXVVEK PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCGTG
      51  GTTTCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
      101 TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
      151 GTGCGGCTGT ACCGTTTTAT GTCGCCGTG GCGTTCGGCG CGGTCGTGTT
      201 CCGCGCGGCG ATACCGTTTG CCGCCGGCTG GTGGGGCAGC GGCTGGGTAC
      251 ACGTCAAACGT GTGTTGGGC TTGATGCTCT TGGCTTACCA GTTGATTGTC
      301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
      30  351 CTGGTACCGC GTGTTCAACG AAATCCCCGT GCTGCTGATG GTTGCCGCGC
      401 TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRIEVD MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWGS GWVHVKLCIG LMLLAYQLYC
      101  GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVEK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFHLLFFVISWFAGLFYLPRIEVDNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPLGF 62
      HP1484  8  FLWVKAFHVIIVISWMAALFYLPRLRFVYHAENAHKKEFVGQVQIEK--KLYSFIASPM 65
      orf102  63  GAVVFGAAIPFAAG---WWGSGVHVHVKLCIGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRNARFY 125
      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
10		10	20	30	40	50	60
	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
15		70	80	90	100	110	120
	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
20		130	140				
	orf102.pep	VFNEIPVLLMVAALYXVVF	KPF	X			
	orf102a	VFNEIPVLLMVAALYL	VVF	KPF	X		
		130	140				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

1	ATGATGTTTT	CTTGGTTCAA	GCTGTTTCAC	TTGTTTTTTG	TCATTCGTG
51	GTTTGCAGGG	CTGTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
25	101	TTGATGTGCC	GCGCGCAAT	CCCAGTATG	TGCGTCTGTC
	151	GTGCGCTGT	ACCGTTTAT	GTCGCCGTTG	GGCTTCGGCG
	201	CGGCGCGCG	ATACCGTTT	CCGCCGGCTG	GTGGGCGAGC
	251	ACGTCAAAC	GTGTTGGGC	TTGATGCTCT	TGGCTTACCA
	301	GGCGTGCTG	TGCGCGGTT	TCAGGATTAC	AGCAATGCTT
30	351	CTGGTACCG	GTGTTCAAC	AAATCCCGT	GCTGCTGATG
	401	TGTATCTGG	CGTGTTCAA	CCGTTTTGA	

This encodes a protein having amino acid sequence <SEQ ID 762>:

1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVN	MAMIDVPRGN	PEYVRLSGMA
35	51	VRLYRFMSPL	GFGAVVFGAA	IPFAAGWWS	GWVHVKLCLG	LMMLLAYQLY
	101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYL	VVF
						PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

40		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVN	MAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL		
		10	20	30	40	50	60
45		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102-1	GFGAVVFGAAIPFAAGWWSGWVHV	KLCLGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
		70	80	90	100	110	120
50		130	140				
	orf102a.pep	VFNEIPVLLMVAALYL	VVF	KPF	X		
	orf102-1	VFNEIPVLLMVAALYL	VVF	KPF	X		
		130	140				

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:

-426-

```

    orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRI FVNMMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
               |||
    orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRI FVNMMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60

5    orf102.pep  GFGAVVFGAAIPFAAGWWSGWVHV KLCGLM LLAYQLYCGVLLRRFQDYSNAF SHRWYR  120
               |||
    orf102ng    GFGAVVFGAAIPFAAGWWSGWVHV KLCGLM LLAYQLYCGVLLRRFQDYSNAF SHRWYR  120

10   orf102.pep  VFNEIPVLLMVAALYXVVF KPF  142
               |||
    orf102ng    VFNEIPVLLMVAALYL VVFKPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1   ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGT
15  51  GTTTCAGGG CTGTTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
    101  TTGATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG
    151  GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGGCG CGGTCGTGTT
    201  CGGCGCGGCG ATACCGTTTG CCGCcggccg GTGGGGCagc ggctggGTTT
    251  ACGTCAAAC TGTTTGGGC TTGATGCTCT TGGCTTATCA GTTGTATTGC
    301  GGCGTGCTGC TGCGCCGTTT TCAGGATFAC AGCAATGCTT TTTCACACCG
20  351  CTGGTACGCG GTGTTCAAacg aaATCCCGT GCTGCTGATG GTTGCCGCGC
    401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1   MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDAPRGN PEYVRLSGMA
25  51  VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHV KLCGL LMLLAYQLYC
    101  GVLLRRFQDY SNAF SHRWYR VFNEIPVLLM VAALYL VVFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

10      20      30      40      50      60
orf102-1.pep MMFSWFKLFHLFFVISWFAGLFYLPRI FVNMMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
30  orf102ng  MMFSWFKLFHLFFVISWFAGLFYLPRI FVNMMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
      10      20      30      40      50      60

70      80      90     100     110     120
orf102-1.pep GFGAVVFGAAIPFAAGWWSGWVHV KLCGLM LLAYQLYCGVLLRRFQDYSNAF SHRWYR
35  orf102ng  GFGAVVFGAAIPFAAGWWSGWVHV KLCGLM LLAYQLYCGVLLRRFQDYSNAF SHRWYR
      70      80      90     100     110     120

130     140
orf102-1.pep VFNEIPVLLMVAALYL VVFKPFX
40  orf102ng  VFNEIPVLLMVAALYL VVFKPFX
      130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45  gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)

50  Query: 3   FSWFKLEHLFFVISWFAGLFYLPRI FVNMMAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFYLPRI FV A + V++ +LY F++
    Sbjct: 8   FLWVKAEHVIAVISWMAALFYLPRL FVYHAENAHKKEFVG VVQIQEK--KLYSFIAS PAM 65

55  Query: 63  GAVVFGAAIP-----FAAGRWWSGWVHV KLCGLM LLAYQLYCGVLLRRFQDYSNAF 115
    G + + F +G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALV VLLLAYHFYCKKCMRELEKDPTRRN 121

    Query: 116 HRWYRVFNEIPXXXXXXXXXXXXXFKPF 142
    R+YRVFNE P KPF
60  Sbjct: 122 ARFYRVFNEAPTILMILIVILV VVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1  ATGGCAAAAA TGATGAAATG GCGGGCTGTT GCGGCGGTCTG CGGCGGCAGC
      51  GGGTTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
     101  TTACGGA AAC GTTCAGGCGC GGC // .....
      //.. ATTTCTGTTA CGATTTTGTG CGAACCGGAT ACGCCGATTA AGGCGAAGCT
     101  CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTCGGGC GGTTACAACA
     151  GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
     201  GGTGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGTGACCG
     251  TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTGGG TCGGACGGC
     301  AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
     351  CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGGTC ATCTCCGAAA
     401  TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCG
     451  CCGCGCCGAT AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1  MAKMMKWA AV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
      51  .....
     101  .....
     151  .....
     201  ..... I SFTILSEPDT
     251  PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDGKLATG
     25  301  MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
     351  RDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGPP RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1  ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
      51  ACTCGGGCAA CAGGTTAAAA AGGGCGATT TATTGCGGAA ATCAATTCGA
     101  CCTCGCAGAC CAATACGCTC AATACGAAA AATCCAAGTT GGAAACGTAT
     151  CAGGCGAAGC TGGTGTCCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
     201  ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAAGAGG
     251  ATTTGGAAG CGGCGAGGAT GCGTTTGCCG CCGCCAAAGC CAATGTTGCC
     301  GAGCTGAAGG CTTTAATCAG ACAGAGCAA ATTTCCATCA ATACGCGCA
     35  351  GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
     401  TGGCGATTCT CGTGGAAGAG GGGCAGACTG TGAACGCGGC GCAGTCTACG
     451  CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAATGCA
     501  GATTGCCGAG GGCATATTA CCAAGGTGAA GCGGGGCGAG GATATTTCGT
     551  TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGCGGAA GCTCGACAGC
     601  GTCGACCCCG GGCTGACCAC GATGTCGTCG GCGGTTACA ACAGCAGTAC
     651  GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA
     701  ATCCGGACGG CAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
     751  ATCGACGGCG TGAAAAATGT GCTGATTATT CCGTCGCTGA CCGTGAAAAA
     801  TCGCGGCGGC AAGGCGTTTG TGCGCGTGT GGGTGCAGC GGCAAGGCGG
     851  CGGAACGCGA AATCCGACCG GGTATGAGAG ACAGTATGAA TACCGAAGTA
     901  AAAAGCGGGT TGAAAGAGG GACAAAGTG GTCATCTCCG AAATAACCGC
     951  CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CTTAGGCGGC CCGCCGCGCC
    1001  GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1  ..VSVGAQASGQ IKILYVKLGQ QVKKGDLIAE INSTSQNTNL NTEKSKLETY
      51  QAKLVSAQIA LGSAEKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
     101  ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
     151  PTIVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
     201  VDPGLTTMSS GGYNSSTDAT SNAVYYARS FVPNPDGKLA GMTTQNTVE
     251  IDGVKNVLI PSLTVKNRGG KAFVRVLGAD GKAEREIRT GMRDSMNTVE
     301  KSGLKEGDKV VISEITAAEQ QESGERALGG PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

```

5      orf85.pep      10      20      30      40
      MAKMMKWAAVA AAAAAVWGGWS-LKPEPHVLDITETVRRG
      orf85a          10      20      30      40      50      60
      MAKMMKWAAVA AAAAAVWGGWSY LKPEPQAAYITETVRRGDISRTVSATGEISPSNLVS
                        //
10     orf85.pep      .....ISFTILSEPDTPIKAKLDSVDPGLTTMSSG
      orf85a          210     220     230     240     250     260
      TIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSSG
15     orf85.pep      110     120     130     140     150     160
      GYNSSDTASN A V Y Y Y A R S F V P N P D G K L A T G M T T Q N T V E I D G V K N V L I I P S L T V K N R G G K
      orf85a          270     280     290     300     310     320
      GYNSSDTASN A V Y Y Y A R S F V P N P D G K L A T G M T T Q N T V E I D G V K N V L I I P S L T V K N R G G R
20     orf85.pep      170     180     190     200     210     220
      AFVRVLGADG K A A E R I R T G M R D S M N T E V K S G L K E G D K V V I S E I T A A E Q Q E S G E R A L G G P
      orf85a          330     340     350     360     370     380
      AFVRVLGADG K A A E R I R T G M R D S M N T E V K S G L K E G D K V V I S E I T A A E Q Q E S G E R A L G G P
25     orf85.pep      230
      PRRX
30     orf85a          390
      PRRX

```

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

```

35      1  ATGGCAAAAA TGATGAAATG GGCGGCTGTT GCGGCGGTCG CGGCGGCAGC
      51  GGT TTGGGGC GGATGGTCTT ATCTGAAGCC CGAGCCGACG GCTGCTTATA
      101 TTACGGAAAC GGT CAGGCGC GGCGACATCA GCCGACGGT TTCTGCAACA
      151 GGGGAGATTT CGCCGTCCAA CCTGGTATCG GTCGCGCGCG AGGCATCGGG
      201 GCAGATTAAG AAAC TTTATG TCAAACTCGG GCAACAGGTT AAAAAGGGCG
      251 ATTTGATTGC GGAATCAAT TCGACCTCGC AGACCAATAC GCTCAATACG
40     301 GAAAAATCCA AATTGGA AAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
      351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
      401 AGGATGATGC GACCGCTAAA GAAGATTGTT AAAGCGCACA GGATGCGCTT
      451 GCCGCGGCCA AAGCCAATGT TGCCGAGCTG AAGGCTCTAA TCAGACAGAG
      501 CAAAATTTCC ATCAATACCG CCGAGTCGGA ATTGGGCTAC ACGCGCATT A
45     551 CCGCAACGAT GGACGGCAGC GTGGTGGCGA TTCTCGTGGA AGAGGGGCAG
      601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
      651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
      701 TGAAGGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
      751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
50     801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTACT
      851 ATTATGCCCG TTCGTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
      901 ATGACGACGC AGAATACGGT TGAATCGAC GGTGTGAAAA ATGTGCTGAT
      951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAGGGCG TTTGTGCGCG
      1001 TGTTGGGTGC AGACGGCAAG GCGGCGGAAC GCGAAATCCG GACCGGTATG
55     1051 AGAGACAGTA TGAATACCGA AGTAAAAAGC GGGTTGAAAG AGGGGGACAA
      1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
      1151 GCGCCCTAGG CGGCCCGCCG GCCTGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 770>:

```

60     1  MAKMMKWA AV AAAAAVWG GWSY LKPEPQ AAYITETVRR GDISRTVSAT
      51  GEISPSNLVS VQAQASGQIK KLYVKLGQOV KGD LIAEIN STSQTNTLNT
      101 ESKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATAK EDLESAQDAL
      151 AAKANVAEL KALIRQSKIS INTAESELGY TRITATMDGT VVAILVEEGQ
      201 TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKVKAGQDI SFTILSEPDT
      251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYAR SFV PNPDKLATG
65     301 MTTQNTVEID GVKNVLIIPS LTVKNRGGRA FVRVLGADGK AAEREIRTGM

```

351 RDSMNTEVKSLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30	40	50	60	70	80
	orf85-1	PQAAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQVKKGDILAE					
					10	20	30
10	orf85a.pep	90	100	110	120	130	140
	orf85-1	INSTSQNTNLNTEKSKLETYQAKLVSAQIALGSAEKKYKQAAALWKDDATAKEDLESAQD					
		40	50	60	70	80	90
15	orf85a.pep	150	160	170	180	190	200
	orf85-1	ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST					
20	orf85a.pep	210	220	230	240	250	260
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS					
25	orf85a.pep	270	280	290	300	310	320
	orf85-1	GGYNSSTDASNAVYYARSFVNPDPGKLGATGTTQNTVEIDGVKNVLIIPSLTVKNRGG					
30	orf85a.pep	330	340	350	360	370	380
	orf85-1	RAFVRVLGADGKAAEREIRTGM RDSMNTEVKSLKEGDKVVI SEITAAEQQESGERALGG					
35	orf85a.pep	390					
	orf85-1	KAFVRVLGADGKAAEREIRTGM RDSMNTEVKSLKEGDKVVI SEITAAEQQESGERALGG					
40	orf85a.pep	PPRRX					
	orf85-1	PPRRX					

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

ORF85	1	MAKMMKWAAVA AAAAAVWGGWS.LKPEPHVLDITETVRRG.....	40
ORF85ng	1	MAKMMKWAAVA AAAAAVWGGWSYKPEPQAAYITEAVRRGDISRTVSAT	50
ORF85	ISFTILSEPDT	250
ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDT	250
ORF85	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYARSFVNPDPGKLGATG	300
ORF85ng	251	PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYARSFVNPDPGKLGATG	300
ORF85	301	MTTQNTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM	350
ORF85ng	301	MTTQNTVEIDGVKNVLLIPSLTVKNRGGKAFVRVLGADGKAVEREIRTGM	350
ORF85	152	RDSMNTEVKSLKEGDKVVI SEITAAEQQESGERALGGPPRR	393
ORF85ng	351	KDSMNTEVKSLKEGDKVVI SEITAAEQQESGERALGGPPRR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCAAAAA TGATGAAATG GGC GGCTGTT GCGGCGGTCG CGGCGGCaac
51  GGT TTGGGGC GGATGGTCTT ATCTGAAGCC CGAACCGCAG GCTGCTTATA
101 TTACGGAaac ggTCAGGCGC GGC GATATCA GCCGGACGGT TTCCGCGACG
5   GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAAA AAGCTTTATG TCAAAC TCGG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
10  351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTG GAAAGCGCGCA GGATGCGCTT
451 GCCCGCGGCC AAGCCAATGT TGCCGAGTTG AAGCTTTAA TCAGACAGAG
501 CAAAATTCCC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATT
551 CCGCGACGAT GGACGGCACG GTGGTGCGGA TTCCCGTGGA AGAGGGGCG
601 ACTGTGAACG CGGCGCAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
15  651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGCGCAT ATTACCAAGG
701 TGAAGCGGGG GCAGGATATT TCGTTTACGA TTTGTCCGA ACCGGATACG
751 CCGATTAAAG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCCG TTCGTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
20  901 ATGACGACGC AGAATACGGT TGAAATCGAC GGTGTGAAAA ATGTGTTGCT
951 TATTCCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACCGG
1001 TGT TGGGTGC GGACGGCAAG GCACTGGAAC GCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAAAGC GGGTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCGCCGA GCAGCAGGAA AGCGGCGAAC
25  1151 GCGCCTAGG CGGCCGCCG CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAKMMKWA AVAAAAVWG GWSYLPPEPQ AAYITEA VRR GDISRTVSAT
51  GEISPSNLVS VQAQASGQIK KLYVKLGQOV KKGLIAEIN STTQNTIDM
101 EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
30  151 AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTPT IVQLANLDM LNMQIAEGD ITKVKAGQDI SFTILSEPDT
251 PIKAKLDSVD PGLTMSGG YNSSTDASN AVYYARFV PNPDKLATG
301 MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTVEKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

```

35 ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30      40      50      60      70      80
orf85ng  PQAAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQOVKKGLIAE
40  orf85-1  VSVGAQASGQIKILYVKLGQOVKKGLIAE
                        10      20      30

90      100     110     120     130     140
orf85ng  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45  orf85-1  INSTSQNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
                        40      50      60      70      80      90

150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
50  orf85-1  AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAIVVEEGQTVNAAQST
                        100     110     120     130     140     150

210     220     230     240     250     260
orf85ng  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTMS
55  orf85-1  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTMS
                        160     170     180     190     200     210

270     280     290     300     310     320
orf85ng  GGYNSSTDASNAYYYYARFVFPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60  orf85-1  GGYNSSTDASNAYYYYARFVFPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
                        220     230     240     250     260     270

330     340     350     360     370     380

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMRDSMNTEVKSGLKEGDKVVIS EITAAEQQESGERALGG
              280      290      300      310      320      330

5
orf85ng      390
              PPRRX
              ||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E. coli* membrane fusion protein:

```

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from
membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia
coli] Length = 380

```

Score = 193 bits (485), Expect = 2e-48

Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQVKKGDLIAE 88
           P   Y T  VR GD+ ++V ATG++      V VGAQ SGQ+K L V +G +VKK  L+
Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTL SVAIGDKVKKDQLLG 100

```

```

Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEXXXXXX 148
           I+   N I  ++ L  +A+  A+  L  A  Y RQ  L +  A S++
Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQRLAQTKAVSQDLDLTAAT 160

```

```

Query: 149  XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST 208
           I++++ S++TA+++L YTRI A M G V  I  +GQTV AAQ
Sbjct: 161 EMAVKQAQIGTIDAQIKRNQASLD TAKTNLDYTRIVAPMAGEVTQITTLQGQTVIAAQA 220

```

```

Query: 209  PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS 268
           P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++  V P
Sbjct: 221 PNILTLADMSAMLVKAQVSEADV IHLKPGQKAWFTVLGDPLTRYEGQIKDVLP----- 273

```

```

Query: 269  GGYNSSTDASNAVYYARSFVPNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
           + +  ++A++YYAR VPNP+G L  MT Q  +++ VKNVL IP  + +  G
Sbjct: 274 -----TPEKVND AIFYARFEV PNPNGLLRLDMTAQVHIQLTDVKNVLTIPLSALGDPVG 328

```

```

Query: 329  KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVIS E 372
           +V L  +G+  ERE+  G ++  + E+  GL+  GD+VVI E
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDT DVEIVKLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E. coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A

shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein

45 was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCTCG CAATACCTTG CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAGGCT ATGGATTGT

```


-432-

251 TCACGCTTGC CTGGCAGTTG GCGGCAAATG ACGCGAAACT CCCCCCGGGG
 301 CTGAAAATCA CCAACGGCAA AAAACTTTAT TCCGTCGGCG GTTTGAATAA
 351 GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAACCC GAAGTCGTCA
 401 AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATTT cTTCGCACCG
 5 451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC
 501 CTATACGCTG AAACCTCAAT CGGTGCAGAT CAACGGCCAG GCAGCCAAAC
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR
 10 51 GKLYAEAKFA DGSVTYKGAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
 101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
 151 SLNNIPAQIG YTDDGKTYTL KLKSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
 15 51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
 101 ATTCGGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
 201 TTTCGAGTCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
 251 ATAGAGACAT ACGCAGGGGC AAACGTGTAT CGGAAGCCAA ATTCGCCGAC
 20 301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
 351 CAAGGCTATG GATTGTGTTCA CGCTTGCCTG GCAGTTGGCG GCAAATGACG
 401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
 451 GTCGGCGGTT TGAATAAGGC GGGTACAGGA AAATACAGCA TAGGCGGCGT
 501 GGAAACCGAA GTCGTCAAAT ATCGGGTTCG GCGCGGCGAC GATGCGGTAA
 25 551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
 601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
 651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AAILSAAALPC AYAAGLPQSA VLHYSYSGYGI PATMTFERSG
 30 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
 151 VVGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVQINGQA AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N.*

meningitidis:

				10	20	30
40	orf120.pep			IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSAALPCAYAAGLPXSAVLHYSYSGYGI	PATXXXXXXNAXKIVSTIKVPLYNIRFE			
		10 20 30 40 50 60				
45	orf120.pep	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADG	SVTYGKAGESKTEQSPKAMD	DLFTLAWQL		
	orf120a	SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADG	SVTYGKAXXXXXXQSPKAMD	DLFTLAWQL		
		70 80 90 100 110 120				
50	orf120.pep	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	EVVKYRVRRGDDAVMYFFAP			
	orf120a	AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE	EVVKYRVRRGDDAVMYFFAP			
55		130 140 150 160 170 180				

-433-

```

              160      170      180
orf120.pep    SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
              |||
orf120a       SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
5             190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTAAAAA TATATTTCC GCCGCCATT TGTCCGCCG
51 CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATTCGGCAG CTACGGCATT CCCGCCACNA NNANNTNNGN ACNNNGNGNC
10  151 AATGCTTNCA AAATCGTTTC GACGATTAAG GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GCGGTACGG TTGTCCGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCCGAC
301 GGCACGCTAA CCTACGGCAA AGCGNNNNN ANCNNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTGTGTTCA CGCTTGCNTG GCAGTTGGCG GCAAATGACG
15  401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGGC GGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAAACCGAA CTCGTCAAAT ATCGGGTGCG GCGCGGCGAC GATGCGGTAA
551 TGTATTTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
20  651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSAAALPC AYAAGLPXSA VLHYSYSGYGI PATXXXXXXX
51 NAXKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
25  151 VGGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep    MMKTFKNIFSAAILSAALPCAYAAGLPXSAVLHYSYSGYGI PATXXXXXXXNAXKIVSTIK
30  |||
orf120-1       MMKTFKNIFSAAILSAALPCAYAAGLPQSAVLHYSYSGYGI PATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep    VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADG SVTYGKAXXXXXXQSPKAM
35  |||
orf120-1       VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep    DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
40  |||
orf120-1       DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep    DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
50  |||
orf120-1       DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
              190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep                                IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
      |||
      orf120ng    SAAILSAALPCAYAARLPQSAVLHYSYSGYGI PATMTFERSGNAYKIVSTIKVPLYNIRFE 69

60  orf120.pep    SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAMD LFTLAWQL 90
      |||
      orf120ng    SGGTVVGNTLHPAYYKDIRRGKLYAEAKFADG SVTYGKAGESKTEQSPKAMD LFTLAWQL 129

```

orf120.pep AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDAVMYFFAP 150
 orf120ng AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGDDTVTYFFAP 189
 5 orf120.pep SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 184
 orf120ng SLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKP 223

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

10 1 ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATT TGTCCGCCGC
 51 CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT
 101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTGTA ACGCAGCGGC
 151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
 201 TTTCGAATCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CCTGCCTACT
 15 251 ATAAAGACAT ACGCAGGGGC AAAGTGTATG CGGAAGCCAA ATTCGCCGAC
 301 GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
 351 CAAGGCTATG GATTGTGTTCA CGCTTGCTG GCAGTTGGCG GCAAATGACG
 401 CGAAACTCCC CCCGGGTCTG AAAATCACC ACGGCAAAA ACTTTATTCC
 451 GTCGGCGGCC TGAATAAGGC GGTACGGGA AAATACAGCA TaggCGGCGT
 20 501 GGAAACCGAA GTCGTCAAT ATCGGGTGCG GCGCGCGAC GATACGGTAA
 551 CGTATTCTT CGCACCGTCC CTGAACAATA TTCGGCACA AATCGGCTAT
 601 ACCGACGACG GCAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
 651 CGGACAGGCC GCCAAACGT AA

This encodes a protein having amino acid sequence <SEQ ID 780>:

25 1 MMKTFKNIFS AAILSALPC AYAARLPQSA VLHYSGSYGI PATMTFERSG
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PAYKDIRRG KLYAEAKFAD
 101 GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
 151 VVGLNKAGTG KYSIGGVETE VVKYRVRRGD DTVTYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVQINGQA AKP*

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

35 orf120-1.pep 10 20 30 40 50 60
 MMKTFKNIFS AAILSALPCAYAAGLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
 orf120ng MMKTFKNIFS AAILSALPCAYAARLPQSAVLHYSGSYGIPATMTFERSGNAYKIVSTIK
 40 orf120-1.pep 70 80 90 100 110 120
 VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
 orf120ng VPLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
 45 orf120-1.pep 130 140 150 160 170 180
 DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
 orf120ng DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRRGD
 50 orf120-1.pep 190 200 210 220
 DAVMYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX
 orf120ng DTVTYFFAPSLNNIPAQIGYTDDGKTYTLKLKSVQINGQAAKPX

This analysis, including the presence of a putative leader sequence in the gonococcal protein
 55 suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

-435-

```

1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  GTGCCGGTGC
51  .GCGTTTGCC  GCCTTGGTCT  GGCTGGTTTT  CGCGCTCGGC  GATACTTTGA
101 CTCCGTTTGC  GGTTCGGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
5  151  GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
201  GATGGTGTTC  TCCTTGATTT  TGTGTGTGGC  ATTATTGTTG  ATTATCGTCC
251  CTATGCTGGT  CGGGCAGTTC  AACAAATTGG  CATCGCGCCT  GCCCAATTA
301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAAA  ATACAATCGG
351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
10  401  ATACGGGAGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
451  AGGCAGGGCG  GCAATATT..

```

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

```

1  MYRRKGRGIK  PWMGAGXAFA  ALVWLVFALG  DTLTPFAVAA  VLAYVLDPLV
51  EWLQKKGLNR  ASASMSVMVF  SLILLALLLL  IIVPMLVGQF  NNLASRLPQL
15  101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
151  RQGGNI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

```

1  ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  GTGCCGGTGC
51  GCGTTTGCC  GCCTTGGTCT  GGCTGGTTTT  CGCGCTCGGC  GATACTTTGA
20  101  CTCCGTTTGC  GGTTCGGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
151  GAATGGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGTCTGT
201  GATGGTGTTC  TCCTTGATTT  TGTGTGTGGC  ATTATTGTTG  ATTATCGTCC
251  CTATGCTGGT  CGGGCAGTTC  AACAAATTGG  CATCGCGCCT  GCCCAATTA
301  ATCGGTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAAA  ATACAATCGG
25  351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAGGCGC
401  ATACGGGAGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
451  AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
501  CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGCGGCA
551  TTGCCAAACT  GGTTCGAGG  CGTTTTGCCG  GTGCTTATAC  GCGCATTACA
601  GGCAATTTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
30  651  AATGCTGATT  ATGGGCTTGG  TTTACGTTT  GGGATTGGTG  CTGGTCGGGC
701  TGGATTCGGG  GTTGGCCATC  GGTATGCTTG  CCGGTATTTT  GGTGTTTGTC
751  CCTTATCTCG  GGGCGTTTAC  GGGATTGCTG  CTTGCCACCG  TCGCCGCCCT
801  GCTCCAGTTC  GGTTCGTGGA  ACGGCATCCT  ATCGGTTTGG  GCGGTTTTTG
851  CCGTAGGACA  GTTTCGCGAA  AGTTTTTTCA  TTACGCCGAA  AATCGTGGA
35  901  GACCGTATCG  GGCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
951  CCGGCAGCTG  ATGGGCTTGG  TCGGAATGTT  GCGGGGATTG  CCTTTGCCCG
1001  CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  TGCAGAAATA  TTTTGCCGGC
1051  AGTTTTTACC  GGGGCAGGTA  G

```

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

```

40  1  MYRRKGRGIK  PWMGAGXAFA  ALVWLVFALG  DTLTPFAVAA  VLAYVLDPLV
51  EWLQKKGLNR  ASASMSVMVF  SLILLALLLL  IIVPMLVGQF  NNLASRLPQL
101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELSN  ALKAWFPVLM
151  RQGGNISSSI  GNLLLLPLLL  YFLLDWQRW  SCGIKLVPR  RFAGAYTRIT
45  201  GNLNEVLGEF  LRQQLVMLI  MGLVYGLGLV  LVGLDSGFAI  GMLAGILVFV
251  PYLGAFTGLL  LATVAALLQF  GSWNGILSVW  AVFAVGQFLE  SFFITPKIVG
301  DRIGLSPFWV  IFSLMAFQQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
351  SFYRGR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N.*

meningitidis:

```

10      20      30      40      50      60
orf121.pep  MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR
55  orf121a    MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAALVAYVLDPLVEWLQKKGLNR
10      20      30      40      50      60
orf121.pep  ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
70      80      90      100     110     120

```

-436-

```
orfl21a      |||||
              ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV
              70      80      90      100      110      120

5
              130      140      150
orfl21.pep    EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
              |||||
orfl21a      EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180

10
orfl21a      SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
              190      200      210      220      230      240
```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```
15
1   ATGTATCGGA  GGAAAGGGCG  GGGCATCAAG  CCGTGGATGG  ATGCCGGTGC
51   GGCCTTTGCC  GCCTTGGTCT  GGCTGGTTTT  CGCGCTCGGC  GATACTTTGA
101  CTCCGTTTGC  GGTTCGGCG  GTGCTGGCGT  ATGTATTGGA  CCCTTTGGTC
151  GAATCGTTGC  AGAAAAAGGG  TTTGAACCGT  GCATCCGCTT  CGATGCTGTG
201  GATGGTGTTC  TCCTTGATTT  TGTGTGTGGC  ATTATTGTTG  ATTATTGTCC
251  CTATGCTGGT  CGGGCAGTTC  AACAAATTGG  CATCGCGCCT  GCCCAATTA
301  CTTGCTTTTA  TGCAGAACAC  GCTGCTGCCG  TGGTTGAAA  ATACAATCGG
351  CGGATATGTG  GAAATCGATC  AGGCATCTAT  TATTGCGTGG  CTTCAAGGCGC
401  ATACGGGCGA  GTTGAGCAAC  GCGCTTAAGG  CGTGGTTTCC  CGTTTTGATG
451  AGGCAGGGCG  GCAATATTGT  CAGCAGTATC  GGCAACCTGC  TGCTGCTTCC
501  CTTGCTGCTT  TACTATTTCC  TGCTGGATTG  GCAGCGGTGG  TCGTGC GGCA
25  551  TTGCCAAACT  GGTTCGAGG  CGTTTGGCCG  GTGCTTATAC  GCGCATTACA
601  GGCAATTTGA  ACGAGGTATT  GGGCGAATTT  TTGCGCGGGC  AGCTTCTGGT
651  GATGCTGATT  ATGGGTTTGG  TTTACGGCTT  GGGGTGGTGG  CTGGTCGGGC
701  TGGATTCGGG  GTTTGCAATC  GGTATGGTTG  CCGGTATTTT  GGTTTTGTG
751  CCCTATTGG  GCGCGTTTAC  AGGACTGCTG  CTGGCAACCG  TCGCCGCCTT
30  801  GCTCCAGTTC  GGTTCGTGGA  ACGGCATCTT  GGCTGTTTGG  GCGGTTTTGG
851  CCGTAGGACA  GTTCTCGAA  AGTTTTCCTA  TTACGCCGAA  AATCGTGGGA
901  GACCGTATCG  GCCTGTCGCC  GTTTTGGGTT  ATCTTTTCGC  TGATGGCGTT
951  CGGGCAGCTG  ATGGGCTTTG  TCGGAATGTT  GGCCGATTG  CCTTTGGCCG
1001 CCGTAACCTT  GGTCTTGCTT  CGCGAGGGCG  TGCAGAAATA  TTTTGCCGGC
35  1051 AGTTTTTACC  GGGGCAGGTA  G
```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```
40
1   MYRRKGRGIK  PWMDAGAAFA  ALVWLVFALG  DTLTPFAVAA  VLAYVLDPLV
51   EWLQKGLNR  ASASMSVMVF  SLILLALLL  IIVPMLVGQF  NNLASRLPQL
101  IGFMQNTLLP  WLKNTIGGYV  EIDQASIIAW  LQAHTGELS  NALKAWFPVLM
151  RQGGNIVSSI  GNLLLLPLLL  YYFLLDWQRW  SCGIAKLVPR  RFAGAYTRIT
201  GNLNEVLGEF  LRQQLLVMLI  MGLVYGLGLV  LVGLDSGFAI  GMVAGLLVVF
251  PYLGAFTGLL  LATVAALLQF  GSWNGILAVV  AVFAVGQFLE  SFFITPKIVG
301  DRIGLSPFWV  IFSLMAFGQL  MGFVGMLAGL  PLAAVTLVLL  REGVQKYFAG
351  SFYRGR*
```

ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```
50
              10      20      30      40      50      60
orfl21a.pep    MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGKLN
              |||||
orfl21-1        MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGKLN
              10      20      30      40      50      60

              70      80      90      100      110      120
orfl21a.pep    ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV
              |||||
55  orfl21-1      ASASMSVMVFSLLLLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLNKNTIGGYV
              70      80      90      100      110      120

              130      140      150      160      170      180
orfl21a.pep    EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              |||||
60  orfl21-1      EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
              130      140      150      160      170      180

              190      200      210      220      230      240
65  orfl21a.pep    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLDSGFAI
```

-437-

	orfl21-1	 SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMGLIMGLVYGLGLVLVGLDSGFAI 190 200 210 220 230 240
5	orfl21a.pep	250 260 270 280 290 300 GMVAGILVFVPYLGAFGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG
10	orfl21-1	250 260 270 280 290 300 GMLAGILVFVPYLGAFGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
	orfl21a.pep	310 320 330 340 350 DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
15	orfl21-1	310 320 330 340 350 DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orfl21.pep	MYRRKGRGIKPWMGAGXAFALVWLVFALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60
	orfl21ng	MYRRKGRGIKPWMGAGAAFAALVWLVYALGDTLTPFAVAAYVLDPLVEWLQKKGLNR	60
25	orfl21.pep	ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21ng	ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orfl21.pep	EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI	156
30	orfl21ng	EIDQASIIAWFQAHTGELSNAKAWFPVLMKQGGNIVSTIGNLLPPLLLYYFLLDWHRW	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEQ ID 788>:

	1	MYRRKGRGIK	PWMGAGAAFA	ALVWLVYALG	DTLTPFAVAA	VLAYVLDPLV
35	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLLL	IIVPMLVGQF	NNLASRLPQL
	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM
	151	KQGGNIVSTI	GNLLLPPLLL	YYFLDWHRW	SCGIPKLVR	RFAGAYTRIT
	201	GNLNKVGK	LRGQLLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
	251	GGG*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 789>:

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGGCGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTGGA	CCCTTTGGTC
	151	GAATGGTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTGT
45	201	GATGGTGTTT	TCCTTGATT	TGTTGTTGGC	ATTATTGTG	ATTATTGTCC
	251	CTATGCTGGT	CGGCGAGTTC	AATAATTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGGTTTGA	TGCAGAACAC	GCTGCTGCCG	TGTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAAATCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
	401	ATACGGGCGA	GTTGAGCAAC	GCGCTTAAGG	CGTGGTTTCC	CGTTTTGATG
50	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCAGCGGTGG	TCGTGCCGCA
	551	TCGCCAAACT	GGTCCGAGG	CGTTTTCGCG	GTGCTTATAC	GCGCATTACG
	601	GGTAATTTGA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGTC	AGCTTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCCGAC
55	701	TGGATTCGGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTTT	GGTGTTCGTC
	751	CCCTATTGGG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGAGCCCTT
	801	GCTCCAGTTC	GGTTCGTGGA	ACGGAATCTT	GGCTGTTTGG	GCGGTTTTTG
	851	CCGTCGGTCA	GTTTCTCGAA	AGTTTTTCA	TTACGCCGAA	AATGTAGGA
	901	GACCGTATCG	GCCTGTCGCC	GTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGAGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCCGGATTG	CCTTTGGCGG
60	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGCGG	CGCAGAAATA	TTTTGCCGGC
	1051	AGTTTTTACC	GGGCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

	1	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	
	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPOL		
5	101	IGFMQNTLLP	WLKNTIGGYV	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM		
	151	KQGGNIVSSI	GNNLLPPLLL	YYFLLDWQRW	SCGIAKLVPR	RFAGAYTRIT		
	201	GNLNEVLGEF	LRGQLLVMLI	MGLVYGLGLM	LVGLDSGFAL	GMVAGILVFP		
	251	PYLGAFTGLL	LATVAALLQF	GSWNGILAVW	AVFAVGQFLE	SFFITPKIVG		
	301	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGAQKYFAG		
	351	SFYRGR*						
10	ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:							
		10	20	30	40	50	60	
	orf121-1.pep	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
	orf121ng-1	MYRRKGRGIK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
15		10	20	30	40	50	60	
	orf121-1.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPOL	IGFMQNTLLP	WLKNTIGGYV	
	orf121ng-1	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPOL	IGFMQNTLLP	WLKNTIGGYV	
20		70	80	90	100	110	120	
	orf121-1.pep	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM	RQGGNIVSSI	GNLNLPLLY	FLLDWQRW	
	orf121ng-1	EIDQASIIAW	FQAHTGELSN	ALKAWFPVLM	RQGGNIVSSI	GNLNLPLLY	FLLDWQRW	
25		130	140	150	160	170	180	
	orf121-1.pep	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVML	IMGLVYGLG	LVGLDSGFAL	
	orf121ng-1	SCGIAKLVPR	RFAGAYTRIT	GNLNEVLGEF	LRGQLLVML	IMGLVYGLG	LVGLDSGFAL	
30		190	200	210	220	230	240	
	orf121-1.pep	GMLAGILVFP	PYLGAFTGLL	LATVAALLQF	GSWNGILSV	WAVFAVGQF	LESFFITPKIVG	
	orf121ng-1	GMLAGILVFP	PYLGAFTGLL	LATVAALLQF	GSWNGILSV	WAVFAVGQF	LESFFITPKIVG	
35		250	260	270	280	290	300	
	orf121-1.pep	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGVQKYFAG	SFYRGRX	
	orf121ng-1	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGVQKYFAG	SFYRGRX	
40		310	320	330	340	350		
	orf121-1.pep	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGVQKYFAG	SFYRGRX	
	orf121ng-1	DRIGLSPFWV	IFSLMAFGEL	MGFVGMLAGL	PLAAVTLVLL	REGVQKYFAG	SFYRGRX	
45		310	320	330	340	350		

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

	sp P43969 PERM HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349							
	Score = 69.9 bits (168), Expect = 2e-11							
	Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)							
50	Query: 26	VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXXXV	84					
		+Y GD + P +A VL+Y+L+ + +L Q R A++ + VP						
	Sbjct: 32	IYFFGDLIAPLLIALVLSYLLEIPINFLNQYKLCPRMLATILIFGSFIGLAAVFFLVLP	91					
55	Query: 85	MLVGQFNNLASRLPOLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK	143					
		ML Q +L S LP + N WL N Y E ID + + + F + ++ +						
	Sbjct: 92	MLWNQTISSLSDLPAMF----NKSNEWLLNLPKNYPIDYSMVDSIFNSVREKILGFE	147					
	Query: 144	AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXDWQRWSCGIAKLVPRRFAGAYTRITGNL	203					
60		+ + + N+VS D G+++ +P+ A+ R +						
	Sbjct: 148	SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM	206					
	Query: 204	NEVLGEFLRGQXXXXXXXXXXXXXXXXXDSGFAIGMVAGILVFPYXXXXXXXXXXXX	263					
		+ + + ++ G+ + + G+ V VPY						
65	Sbjct: 207	QQQISNYIHGKLEILIVTLITYIIFLIFGLNYPLLLAFVGLSVLVPIYIGAVIVTIPVA	266					

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340
 G+ +PLA + ++
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

1 ..ACTGCTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 51 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 101 TTTGACAGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
 201 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCGTGTC
 20 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 351 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 401 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
 451 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRRECGFLC
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
 151 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGAAG TAGATACCGC
 51 GCCTTTGATT TTTTGGCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTTCGGG TACGAATTCG
 35 151 ACTGCTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT
 201 TTTGTCCTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 251 TTTGACAGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
 351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTGTGG GTTTCGTGTC
 40 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 551 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
 601 GAGCAGCGCG TCGGTAACGG CGTGCAGCAG CGCATCGGCA TCGGAGTGTC
 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAGCTTT
 45 701 CTGCCCTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
 751 CGTCATCGTT TGTGTTCCCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLPLLPKASM KKLMEVPVM PIYSFSGTNS
 50 51 TAFSAAMRLS SSCVIFLSF GKPYQQTAAI LTFECTSCPP RSNAYQQYRR
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAQ IGGDVGTHLR NVRREFGFLC
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQTCRT
 201 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALSDDTV
 251 RHRLCS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N. meningitidis*:

5	orf122.pep				10	20	30
					TAFSAALRLSPSXLVIFLSFGKPYQQTAAI		
	orf122a	FLPLLKASMKKLMVEFVPM	PMYSFSGTNS	TAFSAAMRLSSSCVVI	FLSFGKPYQQTAAI		
		30	40	50	60	70	80
10	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR					
	orf122a	LTFFXTSCPPRSNPYQQYRRLRLYAFHAPETEFFVGFAFXVDARNVYAQIGGDVGTHLR					
		90	100	110	120	130	140
15	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT					
	orf122a	NMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT					
20		150	160	170	180	190	200
	orf122.pep	EQRVNGVQQRIGIGVSEQPF	FKWDFNSAKYQ				
25	orf122a	EQRVNGVQQRIGIGVSEQPF	FKWDFNSAKYQLSAFGQLVDIVALS	SDTDV	RHRLCSX		
		210	220	230	240	250	

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

1	ATATCATATT	GGCAAGCAG	TTCAC	TGGAT	TTTTT	GGAAG	TAGATA	ACCGC
51	GCCTTTGATT	TTTTTGCCGC	TCTTAC	CCAA	GGCTTC	GATG	AAAAAG	TTGA
101	TGGTCGAACC	GGTACCGATG	CCGATGT	TATT	CGTTTC	CGGG	TACGAAT	TCG
151	ACTGCNTTTT	CGGCGGCAT	GCGCTTG	AGT	TCGTCTT	TGTG	TCGTCAT	TATT
201	TTTGTCCCTT	GGGAAACCGT	ATCAACA	AAC	AGCCGCC	ATC	TTAACAT	TTTT
251	TTNNNACGTC	CTGCCCGCCG	CGTTCAA	AATC	CTTACC	CAGCA	ATACCG	CCGC
301	CTGCGACTCT	ATGCCTTCCA	TGCGCCC	GAG	ATAACCG	AGT	TTTTCG	TGTG
351	TTTGCCTTT	GANGTTGACG	CACGAA	ATGT	CTATGCC	CAA	ATCGGC	CGCG
401	ATGTTGGCAC	GCATTTGCCG	AATATGC	GGC	GCGAGTT	TGG	GTTTCT	GTGC
451	AATCACGGTC	GTATCGACAT	TGACCG	CC	TG			
501	TTTGATACGC	CGCACGCAA	AGGACG	CGC	TGTCCG	CATC	TTTGA	ACTCT
551	GCGGCGGTGT	CGGGGAAATG	GCTGCC	GATA	TCGCC	CAAAC	CTGCC	GCACC
601	GAGCAGCGCG	TCGGTAACGG	CGTGC	CAGCAG	CGCAT	CGGCA	TCGGAG	TGTC
651	CGAGCAGCCC	TTTTTCAAAT	GGGATTT	CAA	CTCCG	CAAAG	TATCAG	CTTT
701	CTGCCTTCGG	TCAGTTGGTG	GACATCG	TAG	CCCTGT	CCGA	TACGGAT	GTT
751	CGTCATCGTT	TGTGTTCTCTG	A					

This encodes a protein having amino acid sequence <SEQ ID 796>:

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLKASM	KKLMVEP	VPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFXTSCPP	RSNPYQQYRR	
	101	LRLYAFHAP	EITEFFVGFAF	XVDARNVYAQ	IGGDVGTHLR	NMRREFGFLC	
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM	AADIAQTCRT	
50	201	EQRVNGVQQ	RIGIGVSEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALS	SDTDV
	251	RHRLCS*					

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122a.pep	ISYWASSSLDFLEVDTAPLI	FLPLLKASMKKLMVEP	VPM	PMYSFSGTNS	TAFSAAMRLS	
55	orf122-1	ISYWASSSPDFLEVDTAPLI	FLPLLKASMKKLMVEP	VPM	PIYSFSGTNS	TAFSAAMRLS	
		10	20	30	40	50	60
	orf122a.pep	SSCVVIFLSFGKPYQQTAAI	LTFFXTSCPPRSNPYQQYRRLRLYAFHAP	EITEFFVGFAF			
60	orf122-1	SSCVVIFLSFGKPYQQTAAI	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF				

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122-1	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
10	orf122a.pep	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
	orf122-1	FELCGGVGEMAADIAQTCRTEQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
15	orf122a.pep	DIVALSDTDVRHRLCSX					
	orf122-1	DIVALSDTDVRHRLCSX					
		250					
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKKLMVEPVPMYFSFGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFEFTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFEFTSWPPRSNPNYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
35	orf122ng	NVRCEFGFLCNHGRIDIDHLPRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
	orf122.pep	EQRVGNVQQRIGIGVSEQPFFKWFDFNSAKYQ	182
	orf122ng	EQRVGNVQQRVGIRMPEQPFKWFDFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTCCGCCGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTTCGCCAA	GGCTTCGATG	AAGAAATTGa
	101	tgTTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
	151	ACTGCTTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
45	201	TTTAtccttt	gGAAaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
	251	TTTGACAGtc	ctggccgcgc	cggtcaAAc	cgtaccaGca	ataccgcgcg
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATatTGACG	CACGAAATAT	CGataCCCAa	atcggcgcGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
50	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTCT
	551	GCGGCGGTGT	CGGGAATATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggtaaCGG	CGTGACAGCAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
55	701	CTGCCCTTCG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
	751	CGTCATCGTT	TGTGTTCTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFEFTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDLH	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVQQR	RVGIRMPEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSDTDI
	251	RHRLCS*				

-442-

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEFPVPMPIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEFPVPMPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf122-1.pep	SSCVVIFLSFGKPYQQTAAILTFEFTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
	orf122ng	SSCVVIFLSFGKPYQQTAAILTFEFTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122ng	DIDARNIDTQIGGDVGTHLRNVRCEFGFLCNHGRIDIDHPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
		190	200	210	220	230	240
20	orf122-1.pep	FELCGGVGEMAADIAQTCTREQVRVNGVQQRIGIGVSEQPFFKWFDFNSAKYQLSAFGQLV					
	orf122ng	FELCGGVGKMAADVAQTCTREQVRVNGVQQRVGIRMPEQFFKWFDFNSAKYQLSAFGQLV					
		190	200	210	220	230	240
25		250					
	orf122-1.pep	DIVALSDTDVRHRLCSX					
	orf122ng	DIVALSDTDIRHRLCSX					
30		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGCGCGCA GTGCGAACAA CATTTCGCGC CGTTTTCGCG AAACACCCGT
	51	CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG
	101	TTACCGAATA TGAAACTTC CTGCTGCTTA TCGGCTCGGT ATTTGCCCGC
	151	ATGGGCGCGA TTTTGATTGC CGACTTTTC GTCTTGAAAC GCGGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA RFAETPVAVS VTLIGTVLAV MLPVTEYENF LLLIGSVFAP
	51	MGGFDCRLFR LETA*

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTGCGGCA ATGCTCCTC TCCTTCATCT TCCTCCGCCA TCGGGCTGAT
	51	TTGGTTCCGC GCGCGGTAT CGATTGCCGA AATCAGCAGC GGTACGCTGC
45	101	TTGCGCCTTT GGGCTGGCAG CGCGGTCTGG CGGCTCTACT TTTGGGTCAT
	151	GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
	201	CGGACGCAGC TCGATGGAAG GCGTGCGCCT GTCGTTCCGC AAACGCGGTT
	251	CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
	301	GTGATGATT ACGCCGGCGC AACGCTCAGC TCCGCTTTGG GCAAAGTGTT
50	351	GTGGGACGGC GAATCTTTTG TCTGGTGGGC ATTGGCAAAC GCGCGCTGA
	401	TTGTGCTGTG GCTGGTTTTC GCGCACGCA AAACAGGCGG GCTGAAAACC
	451	GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAAGT
	501	CTTTTCCACG GCAGGCAGCA CCGCCGCACA GGTTCAGAC GGCATGAGTT
	551	TCCGAACGGC AGTCGAGCTG TCCGCCGTGA TGCCGCTTTC CTGGCTGCCG
55	601	CTTGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCCT
	651	GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
	701	GTTTGGCAGC GCGTGTGTT ACCGGAGAAA CCGACGTGGC AAAATCCTG
	751	CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTCTCCAC

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

20 Computer analysis of this amino acid sequence gave the following results:

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SAAIGLIWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPPAATLTAT LAYTLTGCMW YALGLAAALF TGETDVAKIL
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPIAVAVAV
 5 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSA	AGLIWFGAAVS	IAEISTGTLLAP	LGWQRLAALLL	GHAVGGALFFAA	
	orf125-1	MSGNASSPSSSA	AGLIWFGAAVS	IAEISTGTLLAP	LGWQRLAALLL	GHAVGGALFFAA	
		10	20	30	40	50	60
15	orf125a.pep	70	80	90	100	110	120
	orf125-1	70	80	90	100	110	120
20	orf125a.pep	130	140	150	160	170	180
	orf125-1	130	140	150	160	170	180
25	orf125a.pep	190	200	210	220	230	240
	orf125-1	190	200	210	220	230	240
30	orf125a.pep	250	260	270	280	290	300
	orf125-1	250	260	270	280	290	300
35	orf125a.pep	310	320	330	340		
	orf125-1	310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV	308
50	orf125.pep	MLPVTEYENFLLLIGSVFAPM-GGFDCRLFRLETA	64
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFDCRLFCLETA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

1 MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
 101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAMVPLSWL
 201 PLAADYTRQA RPPAATLTAT TLAYTLTGCMW MYALGLAAAL FTGETDVAKI
 251 LLGAGLGITG ILAVVLSTVT TFLDTYSAG ASANNISARF AEIPVAVGV
 301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCRLFCLETA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
101 TCGCCCCCTT GGGCTGGCAG CGCGTCTGG CGGCCCTGCT TTTGGGTTCAT
5   151  GCCGTCGGCG GCGCGCTGTT TTTTGCGGCG GCGTATATCG GCGCACTGAC
201  CCGACGCAGC TCGATGGAAA GTGTGCGCCT GTCGTTCGGC AAATGCGGTT
251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301  GTGATGATT ACCTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
10  351  GTGGGACGGC GAATCCTTTG TCTGGTGGGC ATTGGCAAAC GCGCACTGA
401  TCGTGCTGTG GCTGTTTTTC GCGCGACGCA GAACGGGCGG GCTGAAAACC
451  GTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
501  GTTCGCTTCG TCCGGCACAA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
551  CCTTCGGAAC GGCAGTCGAA CTGTCCGCGC TCATGCCGCT TTCCTGGCTG
601  CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGGCGCAAC
15  651  CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCCT
701  TGGGTTTGGC GCGCGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
751  CTGTTGGGCG CGGCTTGGG CATACGGGCG ATTCTGGCAG TCGTCTCTC
801  CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGCAG
851  ACAACATTTT CGCGCGTTTT GCGGAAATAC CCGTCGCTGT CGGCGTTACC
20  901  CTGATCGGCA CCGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
951  CTTCTGCTG CTTATCGGCT CGGTATTTGC GCCGATGGCG GCGGTTTTGA
1001 TTGCCGACTT TTTCTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
1051 TTTGCCGGAC TGCTTCTGTG GCTGGCAGGC TTCATCTCTT ACCGCTTCCT
1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
25  1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACC
1201 CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51  AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
101  VMIYVGATVS SALGKVLWDG ESEVWWALAN GALIVLWLVF GARRTGGLKT
151  VSMMLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
201  PLAADYTRQA RRPFAATLTA TLAYTLTGCV MYALGLAAAL FTGETDVAKI
251  LLGAGLGITG ILAVVLSTVT TTFLDYSAG ASANNISARF AEIPVAVGVT
301  LIGTVLAVML PVTEYKNFLL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
35  351  FAGLVWLAG FILYRFLSS GWESSIGLTA PVMSAVAIAT VSVRLFEEKT
401  QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

40  orf125-1.pep 10 20 30 40 50 60
MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
|||||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf125ng-1 MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLLAPLGWQ RGLAALLLGHAVGGALFFAA
10 20 30 40 50 60

45  orf125-1.pep 70 80 90 100 110 120
AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf125ng-1 AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
70 80 90 100 110 120

50  orf125-1.pep 130 140 150 160 170 179
ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLMLLAVLWLSAEVFSSTAGSTAAQ-VS
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf125ng-1 ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLMLLAVLWLSVEVFASSTNAAPAVS
130 140 150 160 170 180

55  orf125-1.pep 180 190 200 210 220 230 239
DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVMYALGLAAAL
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf125ng-1 DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCVMYALGLAAAL
60 190 200 210 220 230 240

65  orf125-1.pep 240 250 260 270 280 290 299
FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDYSAGASANNISARFAETPVAVGVT
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf125ng-1 FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
250 260 270 280 290 300

```

-446-

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVWLWLAG						
5	orf125ng-1	LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIADFFVLKRREEIEGFDFAGLVWLWLAG						
		310	320	330	340	350	360	
		360	370	380	390	400		
10	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFVKKTQSLQRNPSX						
	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFVKKTQSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
20	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
25	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.AC GGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACC GC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCRL	NHTMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGGXTDDEI	VRWRADDIAE	REPQLGGREFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TGCGCGCGC	CATGCTCGCG
	151	CCTGCGGCGG	AAGCGGTGCA	AGCCACGCCC	GAAGTGGTCA	GGCTGGGCAG
40	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACC GC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
45	451	CTCGACGGGC	GGCAAATATT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	CGCTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCCACGTC
	751	TTCGTATATG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCCGCCAG
	801	CGTGCGTTCA	GGGTTGGAAC	TCTGTGTCGC	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCTT	GCGCCCCACG
	901	CTCAACCAAC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGTTT	CATGATCTCC	CCGCGCGTAA
	1001	CCGCCGCCGC	CGCCAGATTG	GCAGTGGCAC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQSI PLWRGIRCL NTHMMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
5  151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
201 WNQSPEHTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRQD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.*

meningitidis:

```

15  orf126.pep  10      20      30      40      50      60
      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAXTVEATP
      |||||
orf126a  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAEVEATP
      10      20      30      40      50      60

20  orf126.pep  70      80      90      100     110     120
      EVVRLGRQSIPLWRGIRCLNTHMMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTTDEI
      |||||
orf126a  EVVRLGRQXIPLWRGIRCHLKTAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
      70      80      90      100     110     120

25  orf126.pep  130     140     150     160     170     180
      VRWRADDIAEREPLGGREFDXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
      |||||
orf126a  VRWRADDIAEREPLGGREFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
      130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
51  ACTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACCTTTC GATAAAGGCT
35  GCCGCCGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
151 CCTCGCGCGG AAGCGGTCGA AGCCACGCCT GAAGTGGTCA GGCTGGGCAG
201 GCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
251 CCATGATGCA NGAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
301 CCTTTATCCA ACGAGTTCGT CCGCATCTC AAACGCGGCG GCGTAGCGGA
40  TGACNAAATC GTCCGTTGGC GCGCCGACGA CATCGCCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAAC
501 GAACGTCCTC TGCCATTGGG AACACGAATG TGCCCCCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
601 TGGAACCAAT CCCCAGANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
45  651 AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
701 TGGTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
801 CGTGCGTTCC GGGCTGGAAC TCTTATCCGC ACTTATGCC GTCCACCCCG
851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCACCGGCTC GCGCCCCACG
50  901 CTCAATCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCCTGAT
951 TGAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
1001 CCGCGCGCGC CGTCAGATTG GCAGTGGCAC TGTGTTGACGG AAAAGANGCG
1051 CCCGAACGCG ATGAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
1101 A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51  PAEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
60  201 WNQSPXXTST LRGIRGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV

```


251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
 301 LNHHNPEIRY NRARRLIEIN GLFRHGMIS PAVTAAAVRL AVALFDGKXA
 351 PERDEESGLA YIRRQD*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5	orf126a.pep	10	20	30	40	50	60
	orf126-1	10	20	30	40	50	60
10	orf126a.pep	70	80	90	100	110	120
	orf126-1	70	80	90	100	110	120
15	orf126a.pep	130	140	150	160	170	180
	orf126-1	130	140	150	160	170	180
20	orf126a.pep	190	200	210	220	230	240
	orf126-1	190	200	210	220	230	240
25	orf126a.pep	250	260	270	280	290	300
	orf126-1	250	260	270	280	290	300
30	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
35	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
40							
45							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAAYVAAAMLAPAAXTVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAAEAVEATP	60
55	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMQENGLIVWHGQDKPLSSEFVRHLKRGGXTDDEI	120
	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMQENGLIVWHGQDKPLSSEFVRHLKRGGVADDEI	120
	orf126.pep	VRWRADDIAEREPQLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE	180
	orf126ng	VRWRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTRQGEHA AAYVAAAML

51 PAEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYGAKTA
 201 WNQSPEHTST LRGIERGEVAR VYTPPEITLNR PVRLHPRYP LYIAPKENHV
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSM STPPSAKPTS SKWRPGLRPT
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC
 51 ATTGCAGCTT GCAGAACAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA
 101 CCCGCCAAGG CGAACACGCC GCCGCCTATG TTGCCGCCGC GATGCTCGCG
 151 CCGCGCGCGG AAGCGGTGCA GGCAACGCC GAAGTCATCA GGCTGGGCAG
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGCTG AACACGCTCA
 251 CGATGATGCA GGAAAACGGC AGCCTGATG TGTGGCACGG GCAGGACAAG
 301 CCATTATCCA GCGAGTTTGT CCGCCATCTC AAACGCGCGG GCGTAGCGGA
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC
 401 AACTCGCGCG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
 451 CTCGACGGGC GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
 501 GAACGTCCTT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG
 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGGCGC GAAAACCGCG
 601 TGGAAACCAAT CCCCCGAGCA CACCAGCACC TTGCGCGGCA TACGCGGCGA
 651 AGTGGCGCGG GTTTACACGC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACACGTC
 751 TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCCGCCAG
 801 CCGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCCCG
 851 CCTTCGCGCA AGCCGACATC CTCGAAATCG CCGCCGGCCT GCGCCCCACG
 901 CTCAACCAAC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCGCCTCAT
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTCC CCGCCCGTAA
 1001 CCGCCGCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
 1051 CCCGAACGTG ATGAAGAAAG CGGTTTGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA
 51 PAEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYGAKTA
 201 WNQSPEHTST LRGIERGEVAR VYTPPEITLNR PVRLHPRYP LYIAPKENHV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

		10	20	30	40	50	60
orf126-1.pep		MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAAYVAAAMLAPAEAVEATP					
45	orf126ng-1	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAAYVAAAMLAPAEAVEATP					
		10	20	30	40	50	60
		70	80	90	100	110	120
50	orf126-1.pep	EVVRLGRQSIPLWRGIRCLNTHMTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
	orf126ng-1	EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI					
		70	80	90	100	110	120
		130	140	150	160	170	180
55	orf126-1.pep	VRWRADDAIEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECVPE					
	orf126ng-1	VRWRADEIAIEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPO					
		130	140	150	160	170	180
60	orf126-1.pep	GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPPEITLNRPVRLHPRYP					
	orf126ng-1	DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPPEITLNRPVRLHPRYP					
		190	200	210	220	230	240
65							

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAIHPAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNPEIRYNRRARRLIEINGLFRHGFMISSPAVTAARLAVALFDGKDAPERDKESGLA					
10	orf126ng-1	LNHHNPEIRYSRERRRLIEINGLFRHGFMISSPAVTAARVLAVALFDGKDAPERDEESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
20	Length = 327
	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
25	Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTRQGEHXXXXXXXXXXXXXXXXXXXXX 62
	RI V G G++G A QL G+++ L ++ G
	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEPV 60
30	Query: 63 IRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
	+ LGR + W + G+L+V G+D F R G DE+
	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
35	Query: 123 WRADEIAEREPQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHECAPQDL 182
	IA EP L GRF ++ E LD RQ L+ALA L++ +
	Sbjct: 114 -----IAALEPDLAGRFRRLFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165
40	Query: 183 QAQYDWWIDCRGYGAKTAWNQSPEHTSLRGIRGEVARVYTPEITLNRPVRLHHPRYPLY 242
	+D V+DC G LRG+RGE+ V T E++L+RPVRLHHPR+P+Y
	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLHHPRHPIY 218
45	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHPAFGEADILEIAAGLRPTLN 302
	I P++ + F++GAT IES+ P + RS +ELL+A YA+HPAFGEA + E AG+RP
	Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAYAMHPAFGEARVTETGAGVRPAPY 278
	Query: 303 HHNPEIRYSRERRRLIEINGLFRHGFMISSP 331
	+ P R ++E R + +NGL+RHGF+++P
	Sbjct: 279 DNLV--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50 819>:

	1	ATGACTGATA	ATCGGGGGTT	TACGCTGGTT	GAATTAATAT	CAGTGGTCTT
	51	GATATTGTCT	GTACTTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCACTGCGGG	CAGCCTTGT	AGAAAATGCA
55	151	CATTTTATGG	AAAAGTTTTA	TCTGCAGAAT	GGGAGGTTTA	AACAAACATC
	201	TACCAAGTGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTGTATCC
	251	GTTTGAATGG	AATCGtCGCG	CGGG..GCTT	TAGACAGTAA	ATTCATGTTG
	301	AAGGCGGTAG	CCATAGATAA	AGATAAAAAT	CCTTTTATTA	TTAAGATGAA
	351	TGAAAATCTA	GTAACCTTTA	ATTGCAAGA	AGTCCGCCAG	TTCGTGTAGT
	401	GACGGGCTGG	ATTATTTTAA	AGGAAATGAT	AAGGACTGCA	AGTTACTTAA
60	451	GTAG				

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
101 KAVAIKDKDN PFIKMENEL VTFICKKSAS CSDGLDYFK GNDKDKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGC GCGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
51 HFMEKFYLN GRFKQSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIKDKKNP FIIKMENELV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N.*

meningitidis:

```

25  orf127.pep  10      20      30      40      50      60
      MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN AHFMEKFYLN
      |||
orf127a  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN AHFMEKFYLN
      |||
      10      20      30      40      50      60
30  orf127.pep  70      80      90      100     110     120
      GRFKQSTKW PSLPIKEAEG FCIRLNGI VARXALDSKFMLK AVAIKDKKNP FIIKMENEL
      |||
orf127a  GRFKQSTKW PSLPIKEAEG FCIRLNGI -ARGALDSKFMLK AVAIKDKKNP FIIKMENEL
      |||
      70      80      90      100     110
35  orf127.pep  130     140     150
      VTFICKKSASS CSDGLDYFKG NDKDKLLKX
      |||
orf127a  VTFICKKSASS CSDGLDYFKG NDKDKLLKX
      |||
40  orf127a  120     130     140     150

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1  ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
45 101 TTGAGAAAGC AAAGATAAAT ACAGTGC GCGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAAT GGGAGATTTA AACAAACATC
201 TACCAAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAGTGACG
50 401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1  MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
51 HFMEKFYLN GRFKQSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIKDKKNP FIIKMENELV TFICKKSASS CSDGLDYFKG NDKDKLLK*

```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
	orf127a.pep	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINTVRAALLEN	AHFMEKFY	LQN
5	orf127-1	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINAVRAALLEN	AHFMEKFY	LQN
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf127a.pep	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGIARGALDS	KFMLKAVAIDK	DKNPFIIKMN	NENLV
	orf127-1	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGIARGALDS	KFMLKAVAIDK	DKNPFIIKMN	NENLV
		70	80	90	100	110	120
		130	140	150			
15	orf127a.pep	TFICKKSASS	CSDDLDFKGN	DKDCKLLKX			
	orf127-1	TFICKKSASS	CSDDLDFKGN	DKDCKLLKX			
		130	140	150			

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

	orf127.pep	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINAVRAALLEN	AHFMEKFY	LQN	60
25	orf127ng	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINAVRAAFLE	NAHFMEKFY	LQN	60
	orf127.pep	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGIARXALDS	KFMLKAVAIDK	DKNPFIIKMN	NENL	120
30	orf127ng	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGI-ARGALDS	KFMLKAVAIDK	DKNPFIIKMN	NENL	119
	orf127.pep	VTFICKKSASS	CSDDLDFKGN	DKDCKLLK				150
	orf127ng	VTFICKKSASS	CSDDLDFKGN	DKDCKLLK				149

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

35	1	ATGACTGATA	ATCGGGGGTT	TACACTGGTT	GAATTAATAT	CAGTGGTCTT	
	51	GATATTGTCT	GTACTTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG	
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGC	GGG	CAGCTTGT	AGAAAATGCA
	151	CATTTTATGG	AAAAGTTT	TCTGCAGAA	T	GGGAGATT	AACAAACATC
	201	TACCAAATGG	CCAAGTTT	GCATTAAAG	A	GGCAGAAG	GGC
40	251	GTTTGAATGG	AATCGCGCGC	GGGGCTTTAG	ACAGTAAAT	CATGTTGAAG	
	301	GCGGTAGCCA	TAGATAAAGA	TAAAAATCCT	TTTATTATTA	AGATGAATGA	
	351	AAATCTAGTA	ACCTTTATTT	GCAAGAAGTC	CGCCAGTTCG	TGTAGTGACG	
	401	GGCTGGATTA	TTTAAAGGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG	

This encodes a protein having amino acid sequence <SEQ ID 826>:

45	1	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSY	RNYVEKAKIN	AVRAAFLENA	
	51	HFMEKFYLN	GRFKQTSTKW	PSLPIKEAEG	FCIRLNGIAR	GALDSKFMLK	
	101	AVAIDKDKNP	FIKMNENLV	TFICKKSASS	CSDDLDFKGN	DKDCKLLK*	

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

		10	20	30	40	50	60
50	orf127-1.pep	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINAVRAALLEN	AHFMEKFY	LQN
	orf127ng-1	MTDNRGFTLV	ELISVVLILSVL	ALIVYPSYRNYVEK	AKINAVRAALLEN	AHFMEKFY	LQN
		10	20	30	40	50	60
		70	80	90	100	110	120
55	orf127-1.pep	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGIARGALDS	KFMLKAVAIDK	DKNPFIIKMN	NENLV
	orf127ng-1	GRFKQTSTKW	PSLPIKEAEGFC	IRLNGIARGALDS	KFMLKAVAIDK	DKNPFIIKMN	NENLV
		70	80	90	100	110	120
60							

-453-

```

                130      140      150
orf127-1.pep    TFICKKSASSCS DGLDYFKGNDKDKCLLKX
                |||
orf127ng-1      TFICKKSASSCS DGLDYFKGNDKDKCLLKX
                130      140      150

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

1  ..GTGTCGCTGG CTTCGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
51  CAACCAAATG CCGAAAACCC GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
101 ATTTATCTGG GGT'TTCAGCA GGGGTATTTC GATT'TGAGTG CCGACGAGAA
15  151 CCCGTA CTG CATATCTGGT CTTTGGCAGT AGAGGAACAG TATTACCTCC
201 TGTATCCCT TTTGCTGATA TTTTGTCTGCA AAAAAACCAA ATCGCTACGG
251 GTGTCGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
301 GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
351 ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTGGCAGG TTCGCTGCTG
20  401 GCGGTTTACG GGCAAACGCA AAACGGCAGA CGGCAACAG CAAATGGAAA
451 ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATGCTT GCCTGCCTGT
501 TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
551 CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
25  601 TCCGACCCGC ATCCTGTGCG CAAGCCCAT CGTATTTGTC GGCAAAATCT
651 CTTATCCCT ATACCTGTAC CATTGGATTT TTATGCTTT GCCTCCGCTC
701 ATTAGAGGCG GGAACAGCT CGGACTGCCT GCCG..

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

1  ..VSLASVIASQ IFLYEDFNQM RKTVELSAVE LSNIYLGFOQ GYFDLSADEN
51  PVLHWSLAV EEQYLLYPL LLIFCKKTK SLRVLRNISI ILFLILTASS
30  101 FLPSGFYTDI LNQNPYYLS TLRFPPELLAG SLLAVYGQTQ NRRQTANGK
151 RQLLSLFCF ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
201 PTRILSASPI VFVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

1  ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
35  51 CGTGCCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCTCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151 GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
201 TTATACCCGC AGGATTAAGC GGATTATATCC TGCCTTTATT GCGGCCGTGT
251 CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTC AAC
40  301 CAAATGCGGA AAACCGTGGA GCTTTCTGCG GTTTTCTTGT CCAATATTTA
351 TCTGGGGTTT CAGCAGGGGT ATTTGATTT GAGTGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
451 CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
501 GCGTAACATC AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCGTTTT
45  551 TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCTA TACTTATTAC
601 CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGTTTCGC TGCTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCCTG CCTGTTCTGT
751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCTGCT TCCTTCCCTG
50  801 CCGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTT GATTTTATTT GCTTTCGCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCTTGCCTG ATCGGCGGTT GCCGCGTTGA
1001 CGGCCCGATT TCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
55  1051 AAACGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTGTGCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAAAATCAT

```

	1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNNRW	LPGGFLGVDI	FFVISGFLIT
	51	<u>GIILSEIQNG</u>	<u>SFSFRDFYTR</u>	<u>RIKRIYPAFI</u>	<u>AAVSLASVIA</u>	<u>SQIFLYEDFN</u>
	101	QMRKTVELSA	VFLSNIYLGF	QQGYFDLSAD	ENPVLHITWSL	AVEEQYLLY
20	151	PLLLIFCCKK	TKSLRVLRLI	<u>SIILFLILTA</u>	<u>SGFLPSGFT</u>	<u>DILNQNTY</u>
	201	LSTLRFPELL	AGSLLAYVQG	TQNGRRQTAN	<u>KQRQLLSLC</u>	<u>FGALLACLFV</u>
	251	<u>IDKHNPFPFG</u>	<u>MTLLLPCLLT</u>	<u>ALLIRSMQYG</u>	<u>TLPTRILSAS</u>	<u>PIVFVGKISY</u>
	301	SLYLYHWIFI	AFAHYITGDK	<u>QLGLPAVSAV</u>	<u>AALTAGSLL</u>	<u>SYYLIEQPLR</u>
	351	KRKMTFFKAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAEANH
25	401	FPETVTLTGD	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LWVWDEKLAN
	451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR
	501	ETVKRIAIAVK	PVYVFANNTS	ISRSPLREEK	LKRFAANQYL	RPIQAMGDIG
	551	KSNQAVFDLI	KDIPNVHWVD	AQKYLPKNTV	EIYGRYLYGD	QDHLTYFGSY
	601	YMGREFHKHE	RIKKSSHGGA	LO*		

Orf128:	1	VSLASVIASQIFLYEDFNQMRKTVLSAVFLSNIIYLGFQQGYFDLSADENPVLHIWLSLAV	60
		++L S IAS IF+Y DFN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWLSLAV	
HI0392:	46	MALVSFIASAIIFIYDNDFNKLKRTIELAIAFLSNFYLGLTQGYFDLSANENPVLHIWLSLAV	105
Orf128:	61	EEQXXXXXXXXXIFCKKTKSLRVLNRNISIILFLILTASSFLPSGFYTDILNQPNNTYYLS	120
		E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS	
HI0392:	106	EGQYYLIFPLILILAYKKFREVKVLFIITLILFFILLATSFVSANFYKEVLHQPNIIYYLS	165
Orf128:	121	TLRFPELLAGSLLVAYGQTQNGRRQTANGKRQLLSSLCFGALLACLFVIDKHNPFIPGMT	180
		LRFPPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T	
HI0392:	166	NLRFPELLVGSLLAIAYHNLSN-KVOLSKOVNNILAILSTLLFSCFLMNNNIAIFIPGIT	224

```

                                     10      20      30
orfl28.pep                        VSLASVIASQIFLYEDFNQMRKTVELSAVF
50 orfl28a                         ILSEIQNGSFSTRDFYTRRIKRIYPAFIAAVSLSIASQIFLYEDFNQMRKTVELSAVF
                                   60      70      80      90     100     110
                                     40      50      60      70      80      90
orfl28.pep                       LSNIYLGFOQGYYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFFCKKTKSLRVLNRNISI
55 orfl28a                       LSNIYLGFOQGYYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFFCKKTKSLRVLNRNISI
                                   120    130    140    150    160    170
                                     100    110    120    130    140    150
60 orfl28.pep                     ILFLILTASSFLPSGFYTDLNQNTYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK

```

-455-

5	orf128a	I L F L I L T A T S F L P S G F Y T D I L N Q P N T Y Y L S T L R F P E L L A G S L L A V Y G Q T Q N G R R Q T A N G K	180	190	200	210	220	230
	orf128.pep	R Q L L S S L C F G A L L A C L F V I D K H N P F I P G M T L L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I	160	170	180	190	200	210
10	orf128a	R Q L L S S L C F G A L L A C L F V I D K H N P F I P G M T L L L P C L L T A L L I R S M Q Y G T L P T R I L S A S P I	240	250	260	270	280	290
	orf128.pep	V F V G K I S Y S L Y L Y H W I F I A F A P L I R G G K Q L G L P A	220	230	240			
15	orf128a	V F V G K I S Y S L Y L Y H W I F I A F A H Y I T G D K Q L G L P A V S A V A A L T A G F S L L S Y Y L I E Q P L R K R	300	310	320	330	340	350
	orf128a	K M T F K K A F F C L Y L A P S L I L V G Y N L Y A R G I L K Q E H L R P L P G A P L A A E N H F P E T V I L T L G D S H	360	370	380	390	400	410

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

1	ATGCAAGCTG	TCCGATACAG	ACCGGAAATT	GACGGATTGC	GGGCCGTCGC
51	CGTGCTATCC	GTCATGATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
101	GATTCCTGGG	GGTGGACATT	TTCTTTGTCA	TCTCAGGATT	CCTCATTACC
151	GGCATCATTC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
201	TTATACCCGC	AGGATTAAGC	GGATTATATC	TGCTTTTATT	GCGGCCGTGT
251	CGCTGGCTTC	GGTGATTGCC	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
301	CAAAATGCGGA	AAACCGTGA	GCTTTCTGCG	GTTTTCTTGT	CCAATATTTA
351	TCTGGGGTTT	CAGCAGGGGT	ATTTGATTTT	GAGTGCCGAC	GAGAACCCCG
401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCCTGTAT
451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
501	GCGTAACATC	AGCATCATCC	TATTTCTGAT	TTTGACTGCC	ACATCGTTTT
551	TGCCAAGCGG	GTTTTATACC	GATATTCTCA	ACCAACCCAA	TACTTATTAC
601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GCAGGTTCCG	TGCTGGCGGT
651	TTACCGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGCAAA	GGAAAACGGC
701	AGTTGCTTTC	ATCACTCTGC	TTCGGCGCAT	TGCTTGCTCG	CCTGTTCTGT
751	ATTGACAAAC	ACAATCCGTT	TATCCCGGGA	ATGACCTGCG	TCCTTCCCTG
801	CCTGCTGACG	GCACTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
901	TCCCTATACC	TGTACCATTC	GATTTTATT	GCTTTCGCCC	ATTACATTAC
951	AGGCGACAAA	CAGCTCGGAC	TGCCTGCCGT	ATCGGCGGTT	GCCGCGTTGA
1001	CGGCCGGAAT	TTCCCTGTTG	AGTTATTATT	TGATTGAACA	GCCGCTTAGA
1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTCT	ATCTCGCCCC
1101	GTCCCTGATA	CTTGTCGGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAAC
1151	AGGAACACCT	CCGCCCGTTG	CCCGGCGCGC	CCCTTGCTGC	GGAAAATCAT
1201	TTTCCGGAAA	CCGTCTGAC	CCTCGGCGAC	TGCAACGCCG	GACACCTGCG
1251	GGGGTTTCTG	GATTATGTCG	GCAGCCGGGA	AGGGTGGAAA	GCCAAAATCC
1301	TGTCCCTCGA	TTCGGAGTGT	TTGGTTTGGG	TAGATGAGAA	GCTGGCAGAC
1351	AATCCGTTAT	GTCGAAAATA	CCGGGATGAA	GTTGAAAAAG	CCGAAGCCGT
1401	TTTTCATTGCC	CAATTCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
1451	GATTTGAAGC	GCAATCCTTC	CTAATACCCG	GGTCCCAGC	CCGATTACAG
1501	GAAACCGTCA	AAAGGATAGC	CGCCGTCAAA	CCCGTCTATG	TTTTTGCAAA
1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAAA	TGAAAAGAT
1601	TTGCCGCAAA	CCAATATCTC	CGCCCCATTC	AGGCTATGGG	CGACATCGGC
1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
1751	GCCGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGTTCTTAT
1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCCTGCTTA	AATCTTCTCG
1851	CGACGGCGCA	TGTCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

1	MQAVRYRPEI	DGLRAVAVLS	VMIFHLNLRW	LPGGFLGVDI	FFVISGFLIT
51	GIILSEIQNG	SFSFRDFYTR	RIKRIYPAFI	AAVSLASVIA	SQIFLYEDFN
101	QMRKTVELSA	VFLSNILYLG	QQGYFDLSAD	ENPVLHIWSL	AVEEQYLLLY
151	PLLLIFCCCK	TKSLRVLRLN	SIILFLILTA	TSFLPSGFYT	DILNQPNYY
201	LSTLRFPELL	AGSLLAVYQ	TQNGRRQTAN	GKRLSSSLC	FGALLACLFV
251	IDKHNPFI	MTLLLPCLLT	ALLIRSMQYG	TLPTIRLSAS	PIVFGKISY
301	SLYLYHWIFI	AFAHYITGDK	QLGLPAVSAV	AALTAGEFSL	SYLLIEQPLR
351	KRKMTFKKAF	FCLYLAPSLI	LVGYNLYARG	ILKQEHRLPL	PGAPLAAENH
401	FPETVLTG	SHAGHLRGFL	DYVGSREGWK	AKILSLDSEC	LWVDEKLAD
451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR

501 ETVKRIAARK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
20	orf128-1	SSFLPSGFYTDILNQNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLSSLC
	orf128a.pep	FGALLACLFVIDKHNPFIIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPISPIVFGKISY
	orf128-1	FGALLACLFVIDKHNPFIIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPISPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
30	orf128a.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILIPGFPAFRFETVKRIAARKPVYVFANNTSISRSPLREEKLKRFAANQYL
40	orf128-1	PVPRFEAQSFILIPGFPAFRFETVKRIAARKPVYVFANNTSISRSPLREEKLKRFAANQYL
	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDIPNVHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N.gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFGQQGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNISI	90
	orf128ng	LSNIYLGFRGLGYFDLSADENPVLHIWISLAVEEQYLLYPLLLIFCYKTKSLRVLNRNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQNTYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGK	232
65	orf128.pep	RQLSSSLCFGALLACLFVIDKHNPFIIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
	orf128ng	RQLSSLLCFGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

orf128.pep VFVGKISYSLYLYHWIFIAFAPLIRGGKQLGLPA 244
|||||
orf128na VFVGKISYSLYLYHWIFIAFAHYITGDKOLGLPAVSAVAALTAGFSLLSYYLIEOPLRKR 352

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

	1	ATGCAAGCTG	TCCGATACAG	GCCTGAAATT	GACGGATTGC	GGGCCGCTCG
	51	CGTGCTATCC	GTCATTATTT	TCCACCTGAA	TAACCGGTGG	CTGCCCGGAG
	101	GATTCTCTGG	GGTGGACATT	TTCTTTTGTA	TCTCGGGATT	CCTCATTACC
10	151	AACATCAATC	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
	251	CCCTGGCTTC	GGTGATTGCT	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAAATGAGGA	AAACCATAGA	GCTTTCTACG	GTTTTTTTGT	CCAATATTTA
	351	TTTGGGGTTC	CGATTGGGGT	ATTTGATTTT	GATTGCGGAC	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCGGTAGAGG	AACAGTATTA	CCTCCTGTAT
15	451	CCTCTTTTGC	TGATATTTCTG	TTACAAAAAA	ACCAAATGCA	TACGGGTGCT
	501	GCGTAATATC	AGCATCATCC	TGTTTTCTGAT	TTTGACCGCA	TCATCGTTTT
	551	TGCCGCGCGG	GTTTTATACC	GACATCCTCA	ACCAACCcaa	TACTATTATC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GTGGGTTCCG	TGTTGGCGGT
20	651	TTACGGGCAA	ACGCAAAACG	GCAGACGGCA	ACAGAAAAAT	GGAAAAACGG
	701	AGTTGCTTTC	ATTACTCTGT	TTCCGGCGCat	tgCTTGTCTG	CGCTTTCGTG
	751	ATCGACAAC	ACGATCCGTT	TATCCCGGGA	ATAACCTGCT	TCTTCTCCGT
	801	CCTGCTGACG	GCGCTGCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
	851	CCCGCATCTC	GTCGGCAAGC	CCCATCGTAT	TTGTCCGCAA	AATCTCTTAT
25	901	TCCCTATACC	TGTACCATTG	GATTTTTTAT	GCCTTCGCC	ATTACATTAC
	951	AGGCGACAAA	CAGTCCGGAC	TGCCTGACCGT	ATCGGCGGTT	GCCGCGTTGA
	1001	CGGCCCGGATT	TTCCCTGTTG	AGCTATTATT	TGATTGAACA	GCCGCTTAGA
	1051	AAACCGGAAGA	TGACCTTCAA	AAAGGCATTT	TTCTGCCTTT	ATCTCGCCCC
	1101	GTCCCTGTATG	CTTGTCCGTT	ACAACCTGTA	TTCAAGAGGG	ATATTGAACC
30	1151	AGGAACACCT	CCGCCCGCTG	CCCGGCACGC	CCGTTGCTGC	GGAAAAATAAT
	1201	TTTCCGGAAA	CCGTCCTGAC	CCTCGGCGAC	TGCAAGCCGC	GACACCTGCG
	1251	GGGGTTTCTG	GATTATGTCTG	GCGGCAGGGA	AGGGTGGAAA	GCTAAAAATC
	1301	TGTCCCTCGA	TTCCGAGTGT	TTGGTTTGGG	TGGATGAGAA	GCTGGCAGAC
	1351	AACCCGTTGT	GCCGAAAATA	CCGGGATGAA	GTTGAAAAAG	CCGAAGCTGT
35	1401	TTTCATTGCC	CAATTCTATG	ATTTGAGGAT	GGGCGGCCAG	CCCGTGCCGA
	1451	GATTTGAAGC	GCAATCCTTC	CTGATACCCG	GCTTCAAGC	CCGATTTCAG
	1501	GAAACGCTCA	AGAGGATAGC	CGCCGTCAAA	CGCTGATATG	TTTTTGCAAA
	1551	CAATACATCA	ATCAGCCGTT	CTCCCTTGAG	GGAGGAAAAA	TTGAAAAGAT
	1601	TGCTCATATA	CCAATACCTC	CGGCCATTTC	GGGCTATGGG	CGACATCGGC
40	1651	AAGAGCAATC	AGGCGGTCTT	TGATTTGGTT	AAAGATATTC	CCAATGTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATACACG
	1751	GACGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATTT	CGGTTCTTAT
	1801	TATATGGGGC	GGGAATTTCA	CAAACACGAA	CGCTTGCTCA	AGCATTTCCC
	1851	AGGCGCGGCA	TTGCAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 834>:

45	1	<u>MQAVRYRPEI</u>	<u>DGLRAVAVL</u>	<u>VIIFHLNNRW</u>	<u>LPGGFLGVDI</u>	<u>FFVISGFLIT</u>
	51	<u>NIILSEIQNG</u>	<u>SFSFRDFYTR</u>	<u>RIKRIYPAFI</u>	<u>AAVSLASVIA</u>	<u>SQIFLYEDFN</u>
	101	<u>QMRKTELST</u>	<u>VFLSNLYLGF</u>	<u>RLGYFDLSAD</u>	<u>ENPVLHITWSL</u>	<u>AVEEQYYLLY</u>
	151	<u>PLLLIFCYKK</u>	<u>TGSLRLVRNI</u>	<u>SIILFLILTA</u>	<u>SKFLPAGFYT</u>	<u>DILNQPCNTY</u>
50	201	<u>LSTLRFPELL</u>	<u>VKSLLAVYGQ</u>	<u>TQNGRRQTEN</u>	<u>GGQRLLSLLC</u>	<u>FGALLVCLFV</u>
	251	<u>IDKHDPPIFG</u>	<u>ITLLLPCLLT</u>	<u>ALLIRSMQYG</u>	<u>TLPTRILSAS</u>	<u>PIVFVGKISY</u>
	301	<u>SLYLYHWIFT</u>	<u>AFAHYITGDK</u>	<u>QLGLPAVSAV</u>	<u>AALTAGFSLL</u>	<u>SYYLIEQPLR</u>
	351	<u>KRKMTFFKFA</u>	<u>FCLYLAPSLM</u>	<u>LVGYNLYSRG</u>	<u>ILKQEHRLPL</u>	<u>PGTPVAEENN</u>
55	401	<u>FPETVLTLDG</u>	<u>SHAGHLRGFL</u>	<u>DYVGGREGWK</u>	<u>AKILSLDSEC</u>	<u>LWVUDEKLAD</u>
	451	<u>NPLCRKYRDE</u>	<u>VEKAEAVFIA</u>	<u>QFYDLRMGGQ</u>	<u>PVPRFEAQSF</u>	<u>LIPGFKARFR</u>
	501	<u>ETVKRIAIAV</u>	<u>PVYVFANNTS</u>	<u>ISRSPLREEK</u>	<u>LKRFAINQYL</u>	<u>RPIRAMGDIG</u>
	551	<u>KSNQAVFDLV</u>	<u>KDIPNVHWVD</u>	<u>AQKYLPKNTV</u>	<u>EIHGRYLYGD</u>	<u>QDHLTYFGSY</u>
	601	<u>YMGREFHKHE</u>	<u>RLLLKHSRGA</u>	<u>LO*</u>		

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

	orf128-1.pep	MQAVRYRPEIDGLRAVAVL SVMIFHLNNRWLPGGFLGVDIFFVISGFLTITGIILSEIQNG : : :
60	orf128ng	MQAVRYRPEIDGLRAVAVLSVII FHLNNRWLPGGFLGV D IFFVISGFL ITNI IL SEIQNG
	orf128-1.pep	SFSFRDFYTRRIKRIYPAFIAA VSLASVI AS QI FL YEDFNQM RKTVELS AVF LSN IYLGF : : :
65	orf128ng	SFSFRDFYTRRIKRIYPAFIAA VSLASVI AS QI FL YEDFNQM RKTIELST VF LSN IYLGF

```

5  orf128-1.pep  QQGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng     RLGYFDLSADENPVLHIWSLAVEEQYYLLYPLLLIFCYKTKTKSLRVLNRNISIILFLILTA

   orf128-1.pep  SSFLPSGFYTDILNQPNNTYYLSTLRFPELLAGSLLAVYGQTQNGRRQTANGKRQLLSSLC
   orf128ng     SSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAVYGQTQNGRRQTENGKRQLLSSLC

10  orf128-1.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY
   orf128ng     FGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASPPIVFGKISY

   orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
   orf128ng     SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF

   orf128-1.pep  FCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAENHFPETVLTGLGDSHAGHLRGFL
   orf128ng     FCLYLAPSLMLVGYNLYSRGILKQEHLRPLPGTPVAENHFPETVLTGLGDSHAGHLRGFL

20  orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng     DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ

   orf128-1.pep  PVPREFEASFLIPGFPARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAANQYL
   orf128ng     PVPREFEASFLIPGFKARFRET VKRIA AVKPVYVFANNTSISRSPLREEKLKRFAINQYL

30  orf128-1.pep  RPIQAMGDIGKSNQAVFDLIKDI PNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
   orf128ng     RPIRAMGDIGKSNQAVFDLVKDI PNVHWVDAQYLPKNTVEIHGRYLYGDQDHLTYFGSY

   orf128-1.pep  YMGREFHKHERLLKSSHGGALQX
   orf128ng     YMGREFHKHERLLKHSRGGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

40  sp|P43993|Y392_HAEIN  HYPOTHETICAL  PROTEIN  HI0392  >gi|1074385|pir||B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)

45  Query: 38  VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1  MDIFFVISGFLITGIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAFIYN 60

50  Query: 98  DFNQMRKTIELSTVFLSNYILGFRLG YFDLSADENPVLHIWSLAVEEQXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I
   Sbjct: 61  DFNKLRTIELAIAFLSNFYLGTLQGYFDLSANENPVLHIWSLAVEGQYYLIFLILILA 120

55  Query: 158 YKTKTKSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNNTYYLSTLRFPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREVKVLFITILILFFILLATSFVSANFYKEVLHQPNIIYYLSNLRFPELLVGSLLAI 180

   Query: 218 YGQTQNGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
60  Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTGTG
5  101  TGGCGTTGGC GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCCGATGCGC
151  GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201  CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTTGGGCG TATGTGTGTT
251  TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10  1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIVIWA YVWFPPFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTC GTTTGACAT TATTTACGAA TACCGCTGGA TGTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
15  101  CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAGAAA
151  GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201  GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251  TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301  TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
20  351  GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401  AGATTTTCCG GCGGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451  GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501  GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTGGCGAGC GAGTTCATCA
551  CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGCGAGTTG
25  601  GCGTATGTC AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651  GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701  GGATATTCCT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

30  1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TLFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101  LVSGEAAIAL RRGYGPIIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151  ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201  AVYQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

40  orf129.pep      10      20      30      40      50
      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      |||
orf129a  MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
      10      20      30      40      50      60

45  orf129.pep      60      70      80
      ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFV
      |||
orf129a  ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPPFVHPSDILVSGEAAIALRRYGPIIAG
      70      80      90      100     110     120

50  orf129a  SLALIANSQA YICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPLAS
      130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTC GTTTGACAT TATTTACGAA TACCGCTGGA TGTTCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

-460-

101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGAAAAA
 151 GCCGGTGGCG CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
 201 GCTGTATGTT ACGCTGTTC GGGGTACGCC GCTGTTGTG CAGATTGTGA
 251 TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
 301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
 401 AGATTTCCTG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
 451 GCGCGTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
 501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTGGCGGAGC GAGTTCATCA
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GGCGGAGTTG
 601 GCGTATGTT AGAATACGAT TACGGGCGCG TATTCGGTTT ATGAAGAACC
 651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
 701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVWAYVWF PFFVHPSDGI
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

orf129a.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 orf129-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 25 orf129a.pep ALRKVSLLYVTLFRGTPLFVQIVWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG
 orf129-1 ALRKVSLLYVTLFRGTPLFVQIVWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGPLIAG
 30 orf129a.pep SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
 orf129-1 SLALIANSQAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
 orf129a.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
 35 orf129-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
 orf129a.pep KRYNPQHRX
 orf129-1 KRYNPQHRX
 40

Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45 orf129.pep IIEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54
 orf129ng MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60
 orf129.pep ALRKVSLLYVTLFRGTPLFVQIVWAYVWF PFFV 88
 50 orf129ng ALRKVSLLYVTLFRGTPLFVQIVWAYVWF PFFVILHTAFLGNAMRQSRRVDPDKGRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55 1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVWAYVWF PFFVILHTAF
 101 LGNAMRQSRR VPDKGRWIAG SLELNCQPRG RKTRGEFFPG ESNLGTPEPN
 151 PLMSGQRRFP GCENWYPPQN FIKK*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGGATTTTc gtTTTGACAT TATTTAcgaA TACCGCTGGA TGTTTCTTTA

```

51 CCGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGGCGGTT
101 CGGtattggG TCTGTTGTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
5 251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTGCGGGAT ACGGGCCGCT
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
10 451 GCGTGTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTGGCGGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GCCGCCCTGA TTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

```

1 MDRFRDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGPLIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
20 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

```

25 orf129-1.pep MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129ng-1 MDRFRDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG
orf129ng-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG

30 orf129-1.pep SLALIANSGAYICEIFRAGIQSIDKGQMEAARSLGLTYPQAMRYVILPQALRRMLPPLAS
orf129ng-1 SLALIANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLAS

35 orf129-1.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
orf129ng-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLE

orf129-1.pep KRYNPQHRX
40 orf129ng-1 KRYNPQHRX

```

In addition, ORF129ng-1 is homologous to an ABC transporter from *A.fulgidus*:

```

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
[Archaeoglobus fulgidus]Length = 224
Score = 132 bits (329), Expect = 2e-30
45 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)

Query: 65 VSLLYVTLFRGTPLFVQIVIWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAGSLAL 124
+S YV + RGTPL VQI+I +F P+ GI + E A G +AL
50 Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99

Query: 125 IANSGAYICEIFRAGIQSIDKGQMEAACSLGLTYPQAMRYVILPQALRRMLPPLASEFIT 184
SGAYI EI RAGI+SI GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI
55 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAARSLGMTYLQAMRYVIFPQAFRNILPALGNEFIA 159

Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTAALIYLLMTTFLGWIFLRLEKR 242
LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+
60 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQK 217

```

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1  ..CTGAAAGAAT GCCGTCTGAA AGACCCTGTT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACTTT
5  101  GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151  CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTC TACGTAAACA
201  cTACGTCCGC ACTTATTACy TGCTCCAAC TTTTGCCGCC GCAGgcTAgT
251  TTGTGGACAG GCGCGGCGw ATTACAAAAC CTGCCCGCyT CCGCGCCCTT
301  GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10  351  TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTCGA CTACCCCAAA
401  CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGCGCGC
451  TTTCTTGGrTG AACGTGAACC CGrTATTTTT CATTACCGTT CCTGCGATTTC
501  TGACCGCCGC CGTATTCGTA CTGTATCTTT TCrCGTTTAT ACCGATATTT
551  CGGGCGAATG CGTTTACAGA CGATCCGGAr TAR

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1  ..LKECRLKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAAPL
101  HLITLGGMMG GVMVWLTA LWHSGFTKLD YPKLCRIAVP ILFAAAVSR
151  FLXNVNPFXX ITVPAILTAA VFVLYLEXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1  ATGCGGCCGT TTTTCGTGCG CGCGCGGTG CTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101  TGGAACCTAT GCTGCCGCGC GCATACGCGC GTTTTTTGAC TCGGCTTTG
25  151  TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCT CGACTTTGAT
201  GCGCGCATTA TTGCTCGCGC CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
251  CTGCTCTGTT TTTTCGTGCG GCCTATTGGC TGGTGTGTCT GCTGTTCTGC
301  GCCCGGCTGA TTTGGCTAGA CCGAAACACC GACAACCTCG CCCTGCTAAT
351  GTTACTTGCC GCGTTCACG TTTTTCAGAC GGCATATGCC GTCAGCGGCG
401  ATTTGAACCT GTTGCGCGCG CAAGTGCATC TAAATATGGC GCGGCTGATG
30  451  TCGTATCCG TCGCGCTCAG TATTCTTTTG GCGCGGAAG CCCTGAAAGA
501  ATGCCGCTCG AAAGACCTGT TTTTATATCC AAATATCGTT TATAAAAACA
551  TCGCCATTAC TTTCCTGCTC TTGCACGCGC CGCCGAACT TTGGCTGCCC
601  GCGCAAACCG CCGGTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651  CAAGCTCGCT GAGCTTCACC ATCACGAACT CTACGTAAA CACTACGTCC
35  701  GCACTTATTA CCTGCTCCAA CTCTTTGCGC CCGCAGGCTA TTTGTGGACA
751  GCGCGCGCGA AATTACAAAA CTGCCCCGCC TCCGCGCCCC TGCACCTGAT
801  TACCCTCGGC GGCATGATGG GCGGCGTGAT GATGGTGTGG CTGACCGCCG
851  GACTGTGGCA CAGCGGCTTT ACCAAACTCG ACTACCCCAA ACTCTGCCGC
901  ATTGCCGTCC CCATCCTTTT CGCCGCGGCC GTCTCGCGCG CTTTCTTGAT
40  951  GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001  CCGTATTCGT ACTGTATCTT TTCACGTTTA TACCGATATT TCGGGCGAAT
1051  GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1  MRPFVFGAAV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
45  51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC
101  ARLIWLDRNT DNFALLMLLA AFTVFQTA YA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAELWLP
201  AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
251  GAAKLQNLPA SAPLHLITLG GMMGGVMMVW LTAGLWHS GF TKLDYPKLCR
50  301  IAVPILFAAA VSRAFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN
351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

-463-

```

                                     10      20      30
orf130.pep                        LKECRLKDPVFIPNIVYKNIAITFLLHAA
                                     |||||
orf130a      LNLRLAQVHLNMAAVMFVSVRSVILLGAEALKECRLKDPVFIPNVVYKNIAITFLLHAA
5              140      150      160      170      180      190

                                     40      50      60      70      80      90
orf130.pep      AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX
10              |||||
orf130a      AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX
              200      210      220      230      240      250

                                     100      110      120      130      140      150
orf130.pep      LQNLPAAPLHLITLGGMMGVMMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
15              |||||
orf130a      LQNLPAAPLHLITLGGMMGSVMVWLTAGLWHSGF TKLDYPKLCRIAVPILFAAAVSRA
              260      270      280      290      300      310

                                     160      170      180      190
orf130.pep      FLXNVNPFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPEX
20              |||||
orf130a      VLMNVNPIFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPEX
              320      330      340      350
```

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

```

25      1  ATGCGGCCGT  TTTTCGTCGG  CGCGGCCGTG  CTGCCATAC  TCGGTGCGCT
      51  GGTGTTTTTC  ATCAACCCCG  GTGCCATCGT  CCTGCACCGC  CAAATTTTCT
     101  TGGAACCTAT  GCTGCCGGCG  GCATACGGCG  GTTTTTTGAC  TCGCGCTTGT
     151  TTGGACTGGA  CGGGTTTTTC  GGTAACCTG  AAACCTGTCG  CGACTTTGAT
     201  GCGCGCATTA  TTGCTCGCCG  CATCCGCTAT  ACTGCCCTTT  TCGCCGCAAA
     251  CTGCCTCGTT  TTTCGTCGCC  GCCTATTGGC  TGGTGTTGCT  GCTGTTCTGC
     301  GCGCGGCTGA  TTTGGCTAGA  CCGAAACACC  GACAACTTCG  CCCTGCTAAT
     351  GTTACTTGCC  GCGTTCCTG  TTTTTCAGAC  GGCATATGCC  GTCAGCGCGC
     401  ATTTGAACCT  GTTGC GCGC  CAAGTGCATC  TAAATATGGC  GGCGGTGATG
     451  TTCGTATCCG  TGCGCGTCAG  TATTCTTTTG  GGCGCGGAAG  CCCTGAAAGA
     501  ATGCCGCTCT  AAAGACCCAG  TATTCATCCC  CAATGTCGTC  TATAAAAACA
     551  TCGCCATTAC  CTTCTGCTC  CTGCACGCG  CCGCCGAAC  TTGGCTGCCT
     601  GCGCAAACCG  CCGGTTTTAC  CTCGCTCGCC  GTCGGCTTTA  TCCTGCTTGC
     651  CAAGCTGCGT  GAGCTTCACC  ATCAGCAACT  CCTGCGCAAA  CACTACGTCC
     701  GCACTTATTA  CCTGCTCCAA  CTCTTTGCCG  CCGCAGGCTA  TTTGTGGACA
     751  GCGCGCGCGA  AATTACAAAA  CCTGCCCGCC  TCCGCGCCCC  TGCACCTGAT
     801  TACCCTCGGT  GGCATGATGG  GCAGCGTGAT  GATGGTGTGG  CTGACTGCCG
     851  GACTGTGGCA  CAGCGGCTTT  ACCAAGCTCG  ACTACCCGAA  ACTCTGCCGC
     901  ATCGCCGTCC  CCATCCTNTT  CGCCGCCGCC  GTTTCGCGCG  CTGTTTTAAT
     951  GAACGTAAAC  CCGATATTCT  TCATACCGGT  CCCC GCAATT  CTGACCGCCG
    1001  CCGTGTTCTG  GCTTTACCTG  CTGACATTCG  TACCGATCTT  TCGGGCGAAC
    1051  GCGTTTACAG  ACGATCCGGA  ATAA
```

This encodes a protein having amino acid sequence <SEQ ID 850>:

```

50      1  MRPFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
      51  LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVL LLLFC
     101  ARLIWLD RNT DN FALLMLLA AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
     151  FVSVRSVILL GAEALKECRL KDPVFIPNVV YKNIAITFLL LHAAAE LWLP
     201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
     251  GAAKLQNLPA SAPLHLITLG GMMGSVMVWV LTAGLWHSGF TKLDYPKLCR
     301  IAVPILFAAA VSRVLMNVN PIFFITVPAI LTAAEVFLYL LTFVPIFRAN
    351  AFTDDPE*
```

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

```

60      orf130a.pep      MRPFVGA AV LAILGALVFF INPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL
      orf130-1          MRPFVGA AV LAILGALVFF INPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL

      orf130a.pep      KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLD RNTDN FALLMLLA
      orf130-1          KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LLLFCARLIWLD RNTDN FALLMLLA

65      orf130a.pep      AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRSVILLGAEALKECRLKDPVFIPNVV
```


10

15

20

25

35

40

45

50

55

60

-465-

```

901 GCCGCTCTCCA TCCTTTTCGC CTCCGCCGCTT TCGCGCGCTG TTTTAATGAA
951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
1001 TGTTCAATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCG
1051 TTTACAGACG ATCCGGAATA A

```

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1 MRPFFVGA AV LAILGALVFF INPGAII LHR QIFLELMLPA AYGGFLTTAL
51 LDRTGFSGNL KPAATLMAVL LLVAAVLLPF LPQLAAFFVA AYWLVLLEFC
101 AWLIWLD RNT DNFALLMLLA AFTVFQ TAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSVLL GTETLKECRL KDPVFIPNVI YKNIAITLLL HAAAEWLPLA
10 QTAGFTALAV GFILLAKLRE LHHHELLRKH YVRTYYLLQL FAAAGYLWTG
201 AAKLQNL PAS APLHLITLGG MTGGVMMVWL TAGLWHS GFT KLDYPKLCRI
251 AVSILFASAV SRAVL MNVN IFFITVPEIL TAAVFMLYLL TFVPIFRANA
301 FTDDPE*
351

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

15 orf130-1.pep MRPFFVGA AVLAILGALVFFINPGAIVLHRQIFLELMLPAAYGGFLTAALLDWTGFSGNL
orf130ng-1 MRPFFVGA AVLAILGALVFFINPGAII LHRQIFLELMLPAAYGGFLTTALLDRTGFSGNL

20 orf130-1.pep KPVATLMAALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFALLMLLA
orf130ng-1 KPAATLMAVLLLVAAVLLPFLPQLAAFFVAAYWLVL LFCAWLIWLD RNTDNFALLMLLA

orf130-1.pep AFTVFQ TAYAVS GDLNLLRAQVHLNMAAVMFVSVRV SILLGAEALKECRLKDPVFIPNIV
25 orf130ng-1 AFTVFQ TAYAVS GDLNLLRAQVHLNMAAVMFVSVRV SVLLGTETLKECRLKDPVFIPNVI

orf130-1.pep YKNIAITFLLLHAAAEWLPLAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ
30 orf130ng-1 YKNIAIT-LLLHAAAEWLPLAQTAGFTALAVGFILLAKLRELHHHELLRKH YVRTYYLLQ

orf130-1.pep LFAAGYLWTGA AKLQNL PASAPLHLITLGGMMGVMMVWL TAGLWHS GFTKLDYPKLCR
orf130ng-1 LFAAGYLWTGA AKLQNL PASAPLHLITLGGMTGGVMMVWL TAGLWHS GFTKLDYPKLCR

35 orf130-1.pep IAVPILFAAAVSRAFLMNVNPIFFITVPAILTAAVFVLYLFTFIPIFRANAFTDDPEX
orf130ng-1 IAVSILFASAVSRAVL MNVNPIFFITVPEILTAAVFMLYLLTFVPIFRANAFTDDPEX

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
45 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATAGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGTA
351 CTGCTTGGA AAG..

```

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

```

5
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATTT TTGGGATATT
151 GGCGGCGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
351 CTGCTTGGA AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

20
25
30
orf131.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orf131a MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED
orf131.pep YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE
orf131a YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
orf131.pep K
orf131a KQGLRRNGLSERVRWX
130

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35
40
1 ATGGAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCATT
51 TACGGTTGCA GGCTGCCGCT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATTT TTGGGATATT
151 GGCGGCGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA
351 TTGTTTGA AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQEGFDCLE KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

50
55
orf131a.pep MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED
orf131-1 MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
orf131a.pep YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
orf131-1 YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
orf131a.pep KQGLRRNGLSERVRWX
|||||

orf131-1 KQGLRRNGLSERVRWX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPLSLED  60

10 orf131.pep  YEIPLSDGNSSVRANEYESAQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
    orf131ng    YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

15 orf131.pep  K 121
    orf131ng    KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

20 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
    51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
    101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```

25 1 ATGGAAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCGATT
    51 TACGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
    101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTTGGGATATT
    151 GCGCGCGAGA GtccgctGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
    201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
    251 ACTTTTATAG GAAAATAGGG AAGTTGAAG CCTGCGGGTT GGATGGCGT
    301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
    351 CTGTTTGGAA AAGCAGGGGT TCGCGCGCAA CGGCCTGTCC GAGCGCGTCC
    401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

35 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SLSGWCKPR KPAAIDFWDI
    51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
    101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
    orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

    orf131ng-1.pep YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
    orf131-1      YEIPLSDGNRSVRANEYESAQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
    orf131-1      KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
5  51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CTGCCTTAT ATTtCGGCC CGCAATGGCT GTCGGAAGAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
10 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtCGGCG TTTCCGCCCG
451 CCTGCCGCAA ACGCCGCGCC AAGACCCGAA CAGCCAATCG CCGTTTTTcG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTTTCGACAA ACGTTCTAAA
551 TtCGTGCAAT ACCGTCCGCG TACCGCCGTG TTGAACAATC TGGAAATCGA
15 601 CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGACc CAGTTCCACT
651 ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
701 CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGG
751 AAAATTTCGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20 1  MKIHIIGIG GTFMGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDEV EAILNLGLPY ISGPQWLSEN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKERRFRP
151 PAANAAPRPE QPIAVFRHRS RRIHRHLFRQ TFXIRALPSA YRRVEQSGIR
201 PRRHLRLGR DTDPVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
25 251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
30 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CTGCCTTAT ATTtCGGCC CGCAATGGCT GTCGGAAGAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACGC ACGGCAAAAC
35 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATTGGCGGC GTACCGGAAA ATTTCGGCGT TTCGCGCCG
451 CTGCCGCAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTTCGACAAA CGTTCTAAAT
551 TCGTGCAAT CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC
601 CAGGCCGACA TCTTTGCCGA CTGGGCGCG ATACAGACCC AGTTCCACTA
40 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGCTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
801 CTGTTTCGAC GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAATGGG
45 851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCCGT CATTGCCGCC
901 GCGCGTCATG TCGGTGTCGA TATTGAGACC GCCTGCGAAG CCTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAT CAAAGGCACG GCAAACGGCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
1051 ATTCAAGGTT TCGGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACCGCGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
50 1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGCGGCGGTG
1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 GGAAAGCTGC TGGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1  MKIHIIGIG GTFMGLAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMDEV EAILNLGLPY ISGPQWLSEN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
60 151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQOSLQDT LDKGCWTPVE
251 KFGTEHWQA GEANADGSFD VLLDGKTAGR VKWDLMGRHN RMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRRMEIKGT ANGITVYDDF AHHPTAIETT

```

-469-

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV
 401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
 451 GKLEALR*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E.coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQLDEFK 63
 IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
 10 o457: 3 IHIILGICGTFMGGGLAMLARQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-Q 61
 Orf132: 64 ADVVYIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTASML 123
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMA 121
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143
 W+LE G PGF+IGGV G
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf132.pep	MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQLD					
	orf132a	MKHIHIIGIGGTFMGGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAQLD					
		10	20	30	40	50	60
30	orf132.pep	EFKADVYIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA					
	orf132a	EFKADVYIGNVAKRGMVDVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA					
		70	80	90	100	110	120
35	orf132.pep	SMLAWLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPREQPI-----AVFR					
	orf132a	SMLAWLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPREQPI-----AVFR					
		130	140	150	160		
40	orf132.pep	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL					
	orf132a	HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL					
		170	180	190	200	210	220
	orf132a	KRSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVTRVTPSEGLIVCNGRQQSLQD					
		180	190	200	210	220	230

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT
 51 TGCCGCCATT GCCAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCGA
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA
 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC
 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC
 351 GACCACCGCG TCTATGCTCG CGTGGGTTTT GGAATATGCC GGAATCGCAC
 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC
 451 CTGCCGCAAA CGCCGCGCCA AGACCCGAAC AGCCAATCGC CGTTTTTCGT
 501 NTGTAAGCC GACGAATACG ACACCGCGTT TTTGACAAA CGCTCCAAAT
 551 TCGTGCATTA CCGTCCGCGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCCACCA
 651 CCTCGTGGCT ACCGTGCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC
 701 AGCAAAAGCCT GCAAGCACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
 751 AAATTCGGCA CGGAACACGG CTGCCAGGCC GGCGAAGCCA ATGCCGATGG

-470-

5
10
15
20
25
30
35
40
45
50
55

801 CTCGTTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGA
851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC
901 GCGCGTCATG CCGGAGTNGA CATTAGACG GCCTGCGAAG CCTTGAGCAC
951 GTTTAAAAAC GTCAAACGCC GCATGGAAAT CAAAGGCACG GCAAACGGTA
1001 TCACCGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT CGAAACCACG
1051 ATTCAGGTTT TCGGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACCGCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC
1151 CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGGCGCGCGC
1201 GACTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
1301 CAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 ACCAACTGCT TGGACGCTTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

15
20
25
30
35
40
45
50
55

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV
51 YEGFDTAQLD EFKADVYVIG NVAKRGMDV EAILNRGLPY ISGPQWLAEN
101 XLHHHWXLG VAXTHGKTTT SMLAWLEYA GLAPGFIXG VPENFSVSAR
151 LPQTPRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQQSLQDT LDKGCWTPVE
251 KFGTEHGWQA GEANADGSFD VLLDGKKAGH VAWSLMGGHN RMNALAVIAA
301 ARHAGVDIQT ACEALSTFKN VKRMEIKGT ANGITYDDF AHHPTAIETT
351 IQGLRQVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFXYAGGA
401 DWDVAEALAP LGGRLHVGKD FADFVAEIVK NAEAGDHILV MSNGGFGGIH
451 TKLLDALR*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25
30
35
40
45
50
55

orf132a.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
orf132-1 MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD
orf132a.pep EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXLG VAXTHGKTTT
orf132-1 EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVG VAGTHGKTTT
orf132a.pep SMLAWLEYAGLAPGFIXGVPENFSVSARLPQTTPRQDPNSQSPFFVIEADEYDTAFFDK
orf132-1 SMLAWLEYAGLAPGFLIGGVENFGVSARLPQTTPRQDPNSQSPFFVIEADEYDTAFFDK
orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHHLVRTVPSEGLIVC NGRQQSLQDT
orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHYLVRTVPSEGLIVC NGRQQSLQDT
orf132a.pep LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKKAGH VAWSLMGGHNRNALAVIAA
orf132-1 LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNALAVIAA
orf132a.pep ARHAGVDIQTACEALSTFKNVKRMEIKGTANGITYDDFAHHPTAIETT IQGLRQVGG
orf132-1 ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITYDDFAHHPTAIETT IQGLRQVGG
orf132a.pep ARILAVLEPR SNTMKLGTMK AALPASLKEADQVFX YAGGADWDVAEALAP LGGRLHVGKD
orf132-1 ARILAVLEPR SNTMKLGTMK SALPVSLKEADQVFCYAGGVDWDVAEALAP LGGRLNVGKD
orf132a.pep FADFVAEIVK NAEAGDHILV MSNGGFGGIH TKLLDALRX
orf132-1 FADFVAEIVK NAEVGDHILV MSNGGFGGIH GKLLLEALRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60
orf132.pep MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD 60
orf132ng MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQLE 60

-471-

	orf132.pep	EFKADVYVIGNVAKRGMDEVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDEVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLRLGRDTPVPPRAHRTIRRPRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRTLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
20	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDEV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPGKFRFRFP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRLGR	DTDPVPPRA	HRTIRRPRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRTLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCACTTGGAA	GAATTTCAAG	CCGATATTTA
	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
30	251	TGAACCGTGG	GCTGCCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGctgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggacGC	ACGGcaaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCCT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGt	gtaccggaAA	ATTTCCGGCT	TTCCGCCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TCGCCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCCTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGCTGGAA
40	751	AAATTTCGGCA	CCGGACACGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAAA	AGCCGGACAC	GTCGCATGGG
	851	ATTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGCTGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTCAAGAC	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTT	GCCCACCACC	CGACCGCCAT	CGAAACCCAG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATCC	TCGCCGTCTT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTGATAACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACCTGC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTFMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TQLEALGIGV
55	51	HEGFDAQAQLE	EFQADIYVIG	NVARRGMDEV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTFRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHHLVR	TVPSEGLIVC	NGQQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWQI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGHGH	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCLRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGFGGGIH
	451	TKLLDALR*				

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAKMYPMPSTQLEALGIGVHEGFDAQAQLE
   orf132-1      MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAKMYPMPSTQLEALGIDVYEGFDAQAQLD

10 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
   orf132-1      EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

15 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1      SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTTPRQDPNSQSPFFVIEADEYDTAFFDK

20 orf132ng-1.pep RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDT
   orf132-1      RSKFVHYRPRTAVLNNLEFDHADIFADLGAIQTQFHYLVVRTVPSEGLIVCNGRQQSLQDT

25 orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGGHNRMNALAVIAA
   orf132-1      LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRMNALAVIAA

30 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1      ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

35 orf132ng-1.pep ARILAVLEPRSNMTMKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRRLRVGKD
   orf132-1      ARILAVLEPRSNMTMKLGTMKSALPVSLKEADQVFCYAGGVWDVAEALAPLGGRNLNVGKD

40 orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1      FDAFVAEIVKNAEVGDHILVMSNGGFGGIHGKLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

40  Query: 22 KEAGFKVSGCDAKMYPMPSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   ++ G +V+G DA +YPPMST LE GI + +G+DA+QLE Q D+ +IGN RG VE
   Sbjct: 21 RQLGHEVTGSDANVYPPMSTILLEKQGIQLIQYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

45  Query: 82 AILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141
   A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV
   Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAVAGTHGKTTTAGMATWILEQCQYKPGFVIGGV 139

50  Query: 142 PENFGVSARLPQTTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
   P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRT +LNNLEFDH
   Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

55  Query: 202 ADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
   ADIF DL AIQ QFHHLVR VP +G I+ +L+ T+ GCW+ E G WQ
   Sbjct: 191 ADIFDDLKAIQKQFHHLVRIIVPGQGRIIWPENDINLKQTMAMGCWSEQLVGEQGHWQAK 250

60  Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGGHNRMNALAVIAAARHAGVDVQTACEALGAFKN 320
   ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+F N
   Sbjct: 251 KLTTDASEWEVLLDGEKVGVEKWSLVGEHNMHNLMAIAAARHVGVPADAANALGSFIN 310

65  Query: 321 VKRREMIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTMKLGM 379
   +RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMTMK+G
   Sbjct: 311 ARRRLELRGEANGVTYVYDDFAHHPTAILATLAALRGKVGGTARIIAVLEPRSNMTMKMGIC 370

Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRRLRVGKDFDTFVAEIVKNARTGDHI 438
K L SL AD+VF W VAE D DT +VK A+ GDHI
Sbjct: 371 KDDLAPSLGRADEVFLQPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

Query: 439 LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L
Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These
10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1  . . CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
15  101  CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
151  ATTAGTGCGG ACTTCGGCGA TTATTTCATG CCGTTCGCCA GCTATTTCGG
201  CACACACCGT ATGCCCAACA TCCAAGAAAT GTATTTTCC CAAATCGGCG
251  ACTCGGCGT TCACACCGCC TTAACCAG AGCGCGCAA CACTTGGCAA
301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAACAG ATGATACATT
20  351  AGGATTAATA CTGGTCGGCT ACCGAGCCG CATCGACAAC TACATCCACA
401  ACGTTTACGG GAAATGGTGG GATTGAACG GGGATATTCC GAGCTGGGTC
451  AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCAWAGACAA
501  AGTCATCAAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
25  551  TTTTCACCAA CCTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCAAC
601  TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
651  CAAACAAGGT TATGGGTGA GCAGGGTTTC CGCCTGCCG CGAGATTACG
701  GACGTTTGA AGTCGGTACG CGCTGGTTGG GCAACAACT GACTTTGGGC
751  GGCGGATGC GCTATTTCCG CAAGAGCATC CGCGGACGG CTGAAGAACG
801  CTATATCGAC GGCACCAACG GGGGAAATAC CAGCAATTC CGGCAACTGG
30  851  GCAAGCGTTC CATCAACAA ACCGAACTC TTGCCGCCA GCCTTTGATT
901  TTWgATTTTa ACGCCGCTTA CGAGCCGAAG AAAAACCTTA TTTTCCGCGC
951  CGAAGTCAAA AATCTGTTTC ACAGCGTTA TATCGATCCG CTCGATGCGG
1001 GCAATGATGC GGCAAC.GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
1051 AAGGACrAG ACGTACCGT TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
1151 CCTTTTgAT GACGATGAGC TACAAGTTTT AA

```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1  . . PGYYGSDDEF KRAFGENSPT XKKHCNRSCG IYEPVLKKYG KKRANNHSVS
51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
40  101  FGEXTYKGL LKQDDTLGLK LVGYRSRIDN YIHNVYKWW DLNGDIPSWV
151  SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
201  FSDASESPNN ASKEDQLKQG YGLSRVSALP R DYGRLEVGT RWLGNKLTIG
251  GAMRYFGKSI RATAERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
301  XDFNAAYEPK KNLI FRAEVK NLFDRRYIDP LDAGNDAAAXE RYYSSFDPKD
45  351  KDXDVT CNAD KTLNGKYGG TSKSVLTNFA RGRTFLMTMS YKF*

```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAAAGAC AAAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
50  101  AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
151  CCCGGTGCCT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
251  TCACGCAGAC CTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT

```

301 TCATCTCAAT TCGGTGCATC TGTCGACAGC AATTTTATTG CCGGACTGGA
 351 TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG
 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT
 451 AATACTACG GCCTGCTGCT AAAAGGTCTG ACCGGCACCA ATTCAACCAA
 501 AGGTAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG
 551 CATCTGTCGG TGTGCTTTAC GGGCAGAGCA GCGCGAGCGT GCGCGAAAAT
 601 TACCGCGTGG GCGGCGGCGG GCAGCACATC GGAAATTTTG GCGCGGAATA
 651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT
 701 TCAATCCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG
 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAAGTAC AaAAATACAT
 801 CGAAGAGCAT GACAAAAGCT GCGGGGAAAA CCTg . CaCCG CAATACGACA
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG
 901 TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTCG
 951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC
 1001 AGTTCAATTA CGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC
 1051 GCAGCTTACA ATTCGGGCGA GCAGAAATAT CCGAAAGGGT CGAAGTTTAC
 1101 AGGCTGGGGG CTTTAAAGG ATTTTGAAAC CTACAACAAC GCGAAAATCC
 1151 TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCGCGA AACCAGTTG
 1201 CAACACCAT TGGGCTTCAA TTATTTCCAC AACGAATACG GCAAAAACCG
 1251 CTTTCTGAA GAATTGGGGC TGTTTTTCGA CGGTCTGAT CAGGACAACG
 1301 GGCTTTATTC CTATTTGGGG CGGTTTAAAG GCGATAAAGG GCTGCTGCCC
 1351 CAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACACGTT
 1401 CTACTTCGAT GCCGCGCTCA AAAAAGACAT TTACCGCTTA AACTACAGCA
 1451 CCAATACCGT CGGCTACCGT TTCGGCGGCG AATATACGGG CTATTACGGC
 1501 TCGGATGACG AATTTAAGCG GGCATTGCGA GAAAACCTCG CGACATACAA
 1551 GAAACATTGC AACCAGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAAT
 1601 ACGGCAAAAA GCGCGCCAAC AACCATTTCG TCAGCATTAG TGCGGACTTC
 1651 GGCATATTAT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC
 1701 CAACATCCAA GAAATGTATT TTTCCCAAAT CCGCGACTCC GCGGTTTACA
 1751 CCGCCTTAAA ACCAGAGCGC GCAAAACACT GGCAATTTGG CTTCAATACC
 1801 TATAAAAAAG GATTGTAAAC ACAAGATGAT ACATTAGGAT TAAACTGGT
 1851 CGGCTACCGC AGCCGCGATC ACAACTACAT CCACAACGTT TACGGGAAAT
 1901 GGTGGGATTT GAACGGGGAT ATTCCGAGCT GGGTCAGCAG CACCGGGCTT
 1951 GCCTACACCA TCCAACATCG CAATTTCAAA GACAAAGTGC ACAAACACGG
 2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT
 2051 CTTACGCCTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC
 2101 GAATCGGCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG
 2151 GTTGAGCAGG GTTTCGCCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG
 2201 GTACGCGCTG GTTGGGCAAC AAAGTACTT TGGGCGGCGC GATGCGCTAT
 2251 TTCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC
 2301 CAACGGGGA AATACAGCA ATTTCCGGCA ACTGGGCAAG CGTTCCATCA
 2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCC
 2401 GCTTACGAGC CGAAGAAAAA CCTTATTTTC CGCGCGAAG TCAAAAATCT
 2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA
 2501 CGCAGCGTTA TTACAGCTCG TTCGACCGA AAGACAAGGA CGAAGACGTA
 2551 ACGTGTATG CTGATAAAAC GTTGTGCAAC GGCAATACG GCGGCACAAG
 2601 CAAAAGCGTA TTGACCAATT TTGACGCGG ACGCACCTTT TTGATGACGA
 2651 TGAGCTACAA GTTTTAA

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50 1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI
 51 PGAFTQDDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFYS TSTDAGRAGG
 101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVQGN
 151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN
 201 YRVGGGQHI GNFGAEYLER RKQRYFVQEG ALKFNSDSGK WERDLQRQQW
 55 251 KYKPYKNYNN QELQKYIEEH DKSWRENLXP QYDITPIDPS SLKQQSAGNL
 301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS LNPYTNLNL
 351 AAYNSGRQKY PKGSKFTGWG LLKDFETYN AKILDNLNTA TFRLPRETEL
 401 QTTLGFNYFH NEYGKNRFEPE ELGLFFDGPD QDNGLYSYLG RFGDKGLLP
 451 QKSTIVQPAG SQYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG
 60 501 SDEFFKRAFG ENSPTYKKHC NRSCGIYEPV LKKYGKKRAN NHSVSIADF
 551 GDYFMPFASY SRTHRPNIQ EMYFSQIGDS GVHTALKPER ANTWFQGFNT
 601 YKKGLLKQDD TLGLKLVGYR SRIDNYIHNV YGKWWDLNGD IPSWVSTGL
 651 AYTIOHRNFK DKVHKHGFEL ELNYDYGRFF TNLAYYQKS TQPTNFSNAS
 701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGTRWLGK KLTLLGAMRY
 751 FGKSIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQETELA RQPLIFDFYA
 801 AYEPPKNLIF RAEVKNLFD RYIDPLDAGN DAATQRYSS FDPKDKDEDV
 851 TCNADKTLN GKYGGTSKSV LTNFARGRTF LMTMSYKF*

Computer analysis of this amino acid sequence gave the following results:

Homology with the probable TonB-dependent receptor HI121 of *H. influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

Orf133: 31 IYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
5 HI121: 563 INEPILHKSGHKAFNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA 622

Orf133: 91 LKPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNHYGKWWDLNGDIPSWV 150
10 HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNHYGVWW--RDGMPTWA 680

Orf133: 151 SSTGLAYTIQHRXFXDKVHXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSASESPNN 210
15 HI121: 681 ESNGFKYTIHQYKPIVKKSGVELEINYDMGRFFANVSAYQRTNQPTNYADASPRPNN 740

Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYID 270
20 HI121: 741 ASQEDILKQGYGLSRVSMLEPKDYGRLELGTWFDQKLTGLAARYYGKSKRATIEEEYIN 800

Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAEPKKNLIFRAEVKNLFDRRYIDP 330
25 HI121: 801 GSR-FKKNTLRRENYAVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

Orf133: 331 LDAGNDAAAXERYYSFDPKDKDXDVTNADKTLNKGKYGGTSKSVLTNFARGRTFLMTMS 390
30 HI121: 860 LDAGNDAAASQRYSSL-----NNSIECAQDSSAC-----GSDKTVLYNFARGRTYILSLN 910

Orf133: 391 YKF 393
35 HI121: 911 YKF 913

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N.*

meningitidis:

```

35 orf133.pep          10      20      30
PGYYGSDDEFKRAFGENSPTXKKHCNRSCTGI
orf133a  FYFDAALKKDIYRLNYSNTVTGYRFGGXYTGYYXSDDEFKRAFGENSPTYXKKHCNQSCGI
450      460      470      480      490      500

40 orf133.pep          40      50      60      70      80      90
YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
orf133a  YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
510      520      530      540      550      560

45 orf133.pep          100     110     120     130     140     150
KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNHYGKWWDLNGDIPSWVS
orf133a  KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDXYIHNHYGKWWDLNGNIPSWVS
570      580      590      600      610      620

50 orf133.pep          160     170     180     190     200     210
STGLAYTIQHRXFXDKVHXXXXXXXXXYDYGRFFTNLSYAYQKSTQPTNFSASESPNNA
orf133a  STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTQPTNFSASESPNNA
630      640      650      660      670      680

55 orf133.pep          220     230     240     250     260     270
SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYIDG
orf133a  SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGLGAMRYFGKSIRATAEERYIDX
690      700      710      720      730      740

60 orf133.pep          280     290     300     310     320     330
TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAEPKKNLIFRAEVKNLFDRRYIDPL
65 orf133a  TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAEPKKNLIFRAEVKNLFDRRYIDPL

```

	1	AAAGACAAAA	AAGTGTTTAC	CGATGCGCGT	GCCGTATCGA	CCCGTCAGGA
	51	TATATTCAAA	TCCANCAGAA	ACCTCGACAA	CATCGTAGCG	ANCATCCCCG
20	101	GTGCGTTTAC	ACANCAANAT	AAAAGCTCGG	GCNTTGTGTC	TTTGAATATT
	151	CGCNGCGACA	CGCGGGTCGG	GCGGGTCAAT	ACNATGGTNG	ACGGCATCAC
	201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GGCGGTTTCAT
	251	CTCAATTCCG	TGCATCTGTC	GACAGCAATT	TTATNCGCCG	ACTGGATGTC
	301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAGGC	CTCAACAGCC	TTGCCGGGTC
25	351	GGCGAATCTG	CGGACTTTAN	GCGTGGATGA	TGTCGTTTCA	GGCAATANTA
	401	CNTACGGCTG	GCTGCTAAAA	GGTCTGACCG	GCACCAATTTC	AACCAAGGTT
	451	AATGCGATGG	CGCGGATAGG	TGCGCGCAAA	TGGCTGGAAA	CGCGAGCATC
	501	TGTCGGTGTG	CTTTACGGGC	ACAGCAGGCG	CAGCGTGGCG	ACAAATTTACC
	551	GCGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAAATATCTG
30	601	GAACGACGCA	AGCAACGATA	TTTTTGAGCAA	GAGGGCGGGT	TGAATTTCAA
	651	TTCCAACAGC	GGAAATFGGG	AGCGGGATTT	CCAAAAGTCG	TACTGCAAAA
	701	CCAAGTG GTA	TCAAAAATAC	GATGCCCCCC	AAGAACTGCA	AAAATACATC
	751	GAAGGTCATG	ATAAAAGCTG	GCGGGAAAAAC	CTGGCGCCGC	AATACGACAT
	801	CACCCCATCT	GATCCGTCCA	GCCTGAAGCN	GCATCTGGCA	GGCAACCTGT
35	851	TTAAATTTGA	ATACGACGGC	GTATTTCAATA	AATACACGGC	GCAATTTTCG
	901	GATTTAAACA	CCAAAATCGG	CAGCCGCAAA	ATCATCAACC	GCAATTTATCA
	951	ATTTCAATTAC	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
	1001	CAGCCTACAA	TTTCGGCAGG	CAGAAAATATC	CGAAAGGGTC	GAATTTTACA
40	1051	GGCTGGGGGC	TTTTNAAAGA	TTTTGAAACC	TACAACAACG	CAAAAAATCCT
	1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
	1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATACGG	CAAAAACCCG
	1201	TTTCCTGAAG	AATTTGGGCT	GTTTTTCGAC	GCTCCGGATC	ANGACAACGG
	1251	GCTTTATTCC	TATTTGGGGC	GGTTTAAGGG	CGATAAAGGG	CTGCTGCCCC
45	1301	AAAAATCAAT	CATTGTCCAA	CCGGCCGGCA	GCCAAATATT	CAACACGGTC
	1351	TACTTCCAGC	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACGAGAC
	1401	CAATACCGTC	GGCTACCGTT	TCGGCGGCNA	ATATACGGGC	TATTACNGCT
	1451	CGGATGACGA	ATTTAAGCGG	GCATTCCGAG	AAAACTCGCC	GACATACANG
	1501	AAACATTGCA	ACCGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAAATA
50	1551	CGGCAAAAAA	CGCGCCAACA	ACCATTCGGT	CAGCATTAGT	CGCGGACTTC
	1601	GCGATTATTT	CATGCCGTTT	GCCAGCTATT	CGCGCACACA	CCGTATGCCC
	1651	AAACATCCAAG	AAATGTATTT	TTCCCAAATC	GCGGACTCCG	GCGTTTCACAC
	1701	CGCCTTAAAA	CCAGAGCGCG	CAAAACCTTG	GCAATTTGGC	TTCAATTACCT
	1751	ATAAAAAAGG	ATTGTTAAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGTC
55	1801	GGCTACCGCA	GCCGCATCGA	CNACTACATC	CACAACGTTT	ACGGGGAATG
	1851	GTGGGATTTG	AACGGGAATA	TTCGAGCTG	GGTCAGCAGC	ACCGGGCTTG
	1901	CCTACACCAT	CCAACACCGC	AATTTCAAAG	ACAAAGTGCA	CAACACGGTT
	1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTCA	CCAACCTTTC
	2001	TTACGCCCTAT	CAAAAAAGCA	CGCAACCCAG	CAACTTCAGC	GATGCGAGCG
60	2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	CACTCAAAAC	AGGTTATTGG
	2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACGTT	TGGAAGTCGG
	2151	TACGCGCTGG	TTGGGCAACA	AACTGACTTT	GGGCGCGCGC	ATGCGCTATT
	2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	CGACGNCACC
	2251	AATGGGGNAN	TNACCAGCAA	TTTCCGGCAA	CTGGGCAAGC	TGTTCCATCAN
65	2301	ACAAACCGAA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTTGT	TNTACGCGG
	2351	CTTACGAGCC	GAAAGAAAAA	CTTATTTTCC	GCGCCGAAGT	CAAAAATCTG
	2401	TTTCGACAGC	GTATATTCGA	TCCGCTCGAT	GCGGGCAATG	ATGCGGCAAC
	2451	GCAGCGTTAT	TACAGTTCGT	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
	2501	CGTGTAAATG	TGATAACACG	TTATGCAACG	GCAAAATACG	CGGCACAAGC
	2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
	2601	GAGCTACAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

1  KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGAFTXQX KSSGXVSLNI
51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
151 NMAAIGARK WLESGASGVV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWKYQKY DAPQELQKYI
251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
301 DLNRTKIGSRK IINRNYQFNY GLSLNPYTNL NLTAAYNSGR QKYPKGSKFT
351 GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
401 FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
451 YFDAALKKDI YRLNYSNTNV GYRFGGXYTG YYXSDDEFKR AFGENSPTYX
501 KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGGLK QDDILGLKLV
601 GYRSRIDXYI HNVYKWWDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
651 FELELNIDYX RFTNLISYAY QKSTQPTNFS DASESPNNAS KEDQLKQYGY
701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
751 NGXXTSNFRQ LGKRSIXQTE TLARQPLIFD XYAAYPEKKX LIFRAEVKNL
801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNT LCNGKYGGTS
851 KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

10      20      30      40
orf133a.pep      KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGAFTXQXKS
25      10      20      30      40      50      60
orf133-1      EAQIQVLEDVHVAKRVPKDKKVFTDARAVSTRQDIFKSSSENLDNIVRSIPGAFTQQDKS
10      20      30      40      50      60
orf133a.pep      SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGGSSQFGASVDSNFXAGLDVVVK
30      70      80      90      100      110      120
orf133-1      SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVVK
10      20      30      40      50      60
orf133a.pep      GSFGSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNMAAIGARKWL
35      70      80      90      100      110      120
orf133-1      GSFGSGSAGINSLAGSANLRTLGVDVVQGNNTYGLLLKGLTGTNSTKGNMAAIGARKWL
10      20      30      40      50      60
orf133a.pep      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
40      70      80      90      100      110      120
orf133-1      ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
10      20      30      40      50      60
orf133a.pep      WERDFQKSYWKTWKYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDPSSLKXQSAGN
45      70      80      90      100      110      120
orf133-1      WERDLQRQWQKYPYKNYNN-QELQKYIEGHDKSWRENLPQYDITPIDPSSLKQXSAGN
10      20      30      40      50      60
orf133a.pep      LFKLEYDGVFNKYTAQFRDLNRTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
55      70      80      90      100      110      120
orf133-1      LFKLEYDGVFNKYTAQFRDLNRTKIGSRKIINRNYQFNYGLSLNPYTNLNLTAAYNSGRQK
10      20      30      40      50      60
orf133a.pep      YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNFYFHNEYGKNRFP
60      70      80      90      100      110      120
orf133-1      YPKGSKFTGWGLLKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNFYFHNEYGKNRFP
10      20      30      40      50      60
orf133a.pep      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
65      70      80      90      100      110      120
orf133-1      EELGLFFDGPQDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LN	YSTNTVGYRFGGX	YTYGXXSDDEFKRA	FGENSP	TYXKHCNQSCGI	YEPVLKKGKRA
5	orf133-1	LN	YSTNTVGYRFGGEY	TGYGSDDEFKRA	FGENSP	TYXKHCNRSCGI	YEPVLKKGKRA
		480	490	500	510	520	530
	orf133a.pep		530	540	550	560	570
	orf133a.pep	NN	HSVSISADFGDY	FMPFASYSR	THRMPN	IQEMYFSQIGD	SGVHTALKPERANTWQFGFN
10	orf133-1	NN	HSVSISADFGDY	FMPFASYSR	THRMPN	IQEMYFSQIGD	SGVHTALKPERANTWQFGFN
		540	550	560	570	580	590
	orf133a.pep		590	600	610	620	630
	orf133a.pep	TY	KKGLLKQDDIL	GLKLVGYR	SRIDXYI	HN	VYGKWWDLNGNIPSWVSSTGLAYTIQHRNF
15	orf133-1	TY	KKGLLKQDDIL	GLKLVGYR	SRIDNYI	HN	VYGKWWDLNGDIPSWVSSTGLAYTIQHRNF
		600	610	620	630	640	650
	orf133a.pep		650	660	670	680	690
	orf133a.pep	KD	KVHKHGFEL	NYDYXRF	TNLSYAYQ	KSTQPTNFSD	ASESPNNASKEDQLKQGYGLS
20	orf133-1	KD	KVHKHGFEL	NYDYG	RFTNLSYAYQ	KSTQPTNFSD	ASESPNNASKEDQLKQGYGLS
		660	670	680	690	700	710
	orf133a.pep		710	720	730	740	750
	orf133a.pep	R	V	SALPRDYGR	LEVGT	RWLG	NKLT
25	orf133-1	R	V	SALPRDYGR	LEVGT	RWLG	NKLT
		720	730	740	750	760	770
	orf133a.pep		770	780	790	800	810
	orf133a.pep	K	R	SIXQTET	LARQPL	I	F
30	orf133-1	K	R	SIXQTET	LARQPL	I	F
		780	790	800	810	820	830
	orf133a.pep		830	840	850	860	870
	orf133a.pep	S	FDPKDKDE	EVT	CNDNT	LCNGKY	G
35	orf133-1	S	FDPKDKDE	DVTC	NADKT	LCNGKY	G
		840	850	860	870	880	

Homology with a predicted ORF from *N.gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPTYXKHCNRSCGI	31
	orf133ng	FYFDAALKKDIYRLNYSTNAINYRFGGEYTYGSENEFKRAFGENSPAYKEHCDPSCGL	560
50	orf133.pep	YEPVLKKGKKRANNHSVSISADFGDYFMPFASYSRTHRMPN	91
	orf133ng	YEPVLKKGKKRANNHSVSISADFGDYFMPFAGYSRTHRMPN	620
55	orf133.pep	KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN	151
	orf133ng	KPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN	680
60	orf133.pep	STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYG	211
	orf133ng	STGLAYTIQHRNFKDKVHKHGFELNYDYG	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGT	271
65	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGT	800
	orf133.pep	TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLI	331
	orf133ng	TNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLI	860

orf133.pep DAGNDAAXERYSSFDPKDKDXDVTCTNADKTLNCGKYGGTSSKSVLTNFARGRTFLMTMSY 391
 orf133ng DAGNDAATQRYSSFDPKDKDEDVTCTNADKTLNCGKYGGTSSKSVLTNFARGRTFLMTMSY 920
 orf133.pep KF 393
 orf133ng KF 922

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a protein having amino acid sequence <SEQ ID 882>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVAKARV
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGGSSQFGAS VDSNFIAGLD
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
 251 LERRKQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLLNLEYD GVFNKYTAQF
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKE
 401 TGWGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNKYGKN
 451 RFPEELGLFF DGPDDQNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
 501 FYFEALKKD IYRLNYSTNA INYREGGEYT GYYGSENEFK RAFGENSPAY
 551 KEHCDPSCGL YEPVLKKYK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
 601 PNIQEMYFSQ IGDSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
 651 VGYRSRIDNY IHNVYKQWD LMGDIPSWVG STGLAYTIRH RNFKDKVHKH
 701 GFLELNYDY GRFTNLNSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
 751 GLSRVSALPR DYGRLEVGT WLGNKLTLLG AMRYFGKSIR ATAERYIDG
 801 TNGNTSNVR QLGRSIRKQT ETLARQPLIF DFYAAEYPPK NLI FRAEVKN
 851 LDFRRYIDPL DAGNDAATQ YSSFDPKDK DEDVTCNADK TLCNGKYGGT
 901 SKSVLTNFAR GRTFLMTMSY KF*

A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

1 ATGAGATCTT CTTTCCGGTT GAAGCCGATT TGTTTTATC TTATGGGTGT
 51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG
 101 AGGCGCAGAT ACAGGTTTTG GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
 151 CCGAAAGACA AAAAAGTGTT TACCGATGCG CGTGCCGTAT CGACCCGTca
 201 gGATGTGTTC AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC
 251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCTTTGAAT
 301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
 351 CACGCAGACC TTTTATTCTG CTTCTACCGA TGCGGGCAGG GCAGGCGGTT
 401 CATCTCAATT CGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
 501 TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA
 551 ATACCTACGG CCTGCTGCTA AAAGGTCTGA CCGGCACCAA TTCAACCAAA
 601 GGTAAATGCG TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
 651 GTCTGTCCGT GTGCTTTACG GGCACAGCAG GCGCGGCGTG GCGCAAAATT
 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAATAT
 751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
 801 CAATGCCGGC AGCGGAAAAT GGGAACGGGA TTTGCAAAGG CAATACTGGA
 851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCCAAGAACT GCAAAAAATC
 901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC
 1001 TGTTTAAATT GGAATACGAC GCGGTATTCA ATAAATACAC GCGCGAATTT
 1051 CGCGATTTAA ACACCAGAAT CGGCAGCCGC AAAATCATCA ACCGCAATTA
 1101 TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTCAATCTGA
 1151 CCGCAGCCTA CAATTGCGGC AGGCAGAAAT ATCCGAAAGG GCGGAAGTTT
 1201 ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACGCGAAAAA
 1251 CCTCGACCTC AACAACACCG CCACCTTCCG GCTGCCCCCG GAAACCGAGT
 1301 TCTCAACACC TTTGGGCTTC AATTATTTCC ACAACGAATA CGGCAAAAAC
 1351 CGCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCTCTG ATCAGGACAA
 1401 CCGGCTTTAT TCCTATTTGG GCGGTTTAA GGGCGATAAA GGGCTGTTGC
 1451 CTCAAAAATC AACCATTGTC CAACCGGCCG GCAGCCAATA TTTCAACACG
 1501 TTCTACTTCG ATGCCGCGCT CAAAAAAGAC ATTTACCGCT TAACTACAG
 1551 CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
 1601 GCTCGGAAAA CGAATTTAAG CGGGCATTTCG GAGAAAACTC GCCGGCATAC
 1651 AAGCAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
 1701 ATACGGCAAA AAGCGCGCCA ACAACCATTTC GGTGAGCATT AGTGCGGACT
 1751 TCGGCGATTA TTTTCATGCCG TTCGCCGGCT ATTCGCGCAC ACACCGTATG

-480-

1801 CCCAACATCC AAGAAATGTA TTTTCCCAA ATCGGCGACT CCGGCGTTCA
 1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA
 1901 CCTATAAAAA AGGATTGTTA AAACAAGATG ATATATTAGG ATTGAAACTG
 1951 GTCGGCTACC GCAGCCGCAT TGACAACACT ATCCACAACG TTTACGGGAA
 5 2001 ATGGTGGGAT TTGAACGGGG ATATTCCGAG CTGGGTCGGC AGCACCGGGC
 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC
 2101 GGTTTTGGAG TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT
 2151 TTCTTACGCC TATCAAAAAA GCACGCAACC GACCAATTTT AGCGATGCGA
 10 2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT
 2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTTGGAAGT
 2301 CGGTACGCGC TGGTTGGGCA ACAAACGTAC TTTGGGCGGC GCGATgcGCT
 2351 ATTTCCGGCA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC
 2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT
 2451 CAAACAAACC GAAACCCCTG CCCGACAGCC TTTGATTTTT GATTTTTTACG
 15 2501 CCGCTTACGA GCCGAAGAAA AACCTTATTT TCCGCGCGCA AGTCAAAAAC
 2551 CTGTTTCGACA GCGGTTATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC
 2601 AACGCAGCGT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG
 2651 TAACGTGTAA TGCTGATAAA ACGTTGTGCA ACGGCAAATA CGGCGGCACA
 2701 AGCAAAAGCG TATTGACCAA TTTCGCACGC GGACGCACCT TCTTGATGAC
 20 2751 GATGAGCTAC AAGTTTTTAA

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAARV
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD
 25 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEY
 251 LERRKQYFV QEGLKFNAG SGKWERDLQR QYWKTKWYK YEDPQELQKY
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQOS AGNLFKLEYD GFVNKYTAQF
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
 401 TGWGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNEYGKN
 451 RFPEELGLFF DGPDDQNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
 501 FYFDAALKKD IYRLNYSNA INYRFGGEYT GYGSSENEFK RAFGENSPAY
 551 KEHCDPSCGL YEPVLKKYK KRNANHVSII SADFGDYFMP FAGYSRTHRM
 601 PNIQEMYFSQ IGDGSHVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
 35 651 VGYRSRIDNY IHNVYKWWDD LNDIPSWVG STGLAYTIRH RNFKDKVHKH
 701 GFELELDYDY GREFTNLSYA YQKSTQPTNF SDASESPNNA SKEDQLKQGY
 751 GLSRVSALPR DYGRLEVGTW WLGKNTLGG AMRYFGKSIR ATAERYIDG
 801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAEPPK NLIFFRAEVKN
 851 LFDRIYIDPL DAGNDAATQR YYSFDPKDK DEDVTCNADK TLCNGKYGGT
 40 901 SKSVLTNFR GRFTLMTMSY KF*

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

10 20 30 40 50 60
 orf133ng-1.pep SFRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVKAARVPKDKKVFTDARAV
 45 orf133-1 EAQIQVLEDVHVKAARVPKDKKVFTDARAV
 10 20 30
 70 80 90 100 110 120
 orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGFRVNTMVDGITQTFYS
 50 orf133-1 STRQDIFKSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGFRVNTMVDGITQTFYS
 40 50 60 70 80 90
 130 140 150 160 170 180
 orf133ng-1.pep TSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
 55 orf133-1 TSTDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
 100 110 120 130 140 150
 190 200 210 220 230 240
 orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI
 60 orf133-1 NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRSVAQNYRVGGGGQHI
 160 170 180 190 200 210
 250 260 270 280 290 300
 orf133ng-1.pep GNFGEEYLERRKQYFVQEGLKFNAGSGKWERDLQRQYWKTKWYKYYEDPQELQKYIEE

-481-

orf133-1
 220 230 240 250 260
 GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQWKYKPYKNYNN-QELQKYIEE

5

orf133ng-1.pep 310 320 330 340 350 360
 HDKSWRENLAPOYDITPIDPSGLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
 orf133-1 HDKSWRENLPQYDITPIDPSSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
 270 280 290 300 310 320

10

orf133ng-1.pep 370 380 390 400 410 420
 NRNYQFNYSLSNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
 orf133-1 NRNYQFNYSLSNPYTNLNLTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDNLNT
 330 340 350 360 370 380

15

orf133ng-1.pep 430 440 450 460 470 480
 ATFRLPRETELQTTLGFNHFHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFKGDKGLL
 orf133-1 ATFRLPRETELQTTLGFNHFHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFKGDKGLL
 390 400 410 420 430 440

20

orf133ng-1.pep 490 500 510 520 530 540
 PQKSTIVQFAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAF
 orf133-1 PQKSTIVQFAGSQYFNTFYFDAALKKDIYRLNYSTNTVGYRFGGEYTGYYGSDDEFKRAF
 450 460 470 480 490 500

25

orf133ng-1.pep 550 560 570 580 590 600
 GENSPAYKEHCDPSCGLYEPVLKKYGGKRRNNHNSVSIADFGDYFMPFAGYSRTHRMPNI
 orf133-1 GENSPYKHKCNRSCTGIYEPVLKKYGGKRRNNHNSVSIADFGDYFMPFASYSRTHRMPNI
 510 520 530 540 550 560

30

orf133ng-1.pep 610 620 630 640 650 660
 QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
 orf133-1 QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNYIHN
 570 580 590 600 610 620

35

orf133ng-1.pep 670 680 690 700 710 720
 VYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
 orf133-1 VYGKWWDLNGDIPSWVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
 630 640 650 660 670 680

40

orf133ng-1.pep 730 740 750 760 770 780
 STQPTNFSDASESNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
 orf133-1 STQPTNFSDASESNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMR
 690 700 710 720 730 740

45

orf133ng-1.pep 790 800 810 820 830 840
 YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLI
 orf133-1 YFGKSIRATAEERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDYAAYEPKKNLI
 750 760 770 780 790 800

50

orf133ng-1.pep 850 860 870 880 890 900
 FRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKS
 orf133-1 FRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDPKDKDEDVTCNADKTLGNGKYGGTSKS
 810 820 830 840 850 860

55

orf133ng-1.pep 910 920
 VLTNFARGRTFLMTMSYKFX
 orf133-1 VLTNFARGRTFLMTMSYKFX
 870 880

60

65

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H. influenzae*:

sp|P45114|YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
>gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913

Score = 930 bits (2377), Expect = 0.0

Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

Query: 38 QVLEDVHVAKARVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSV 88

Query: 98 SLNIRGDSGFGFRVNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFS 157
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAAIDPNFIAGVDVNKSNFS 148

Query: 158 GSAGINSLAGSANLRTLGVDDVVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217
G++GIN+LAGSAN RTLGV+DV+ M RKWL++G
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRKWLNDGG 208

Query: 218 SVGVLYGHSRRGVAQNYRVGGGQHHIGNFGEEYLERRKQYFVQEGGLKFNAGSGKWERD 277
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D
Sbjct: 209 YGVVVYGYSQREVSQDYRI--GGERLASLGQDILAKEKEAYF--RNAGYILNP--EQWTPD 265

Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
L +++W +Y KK +D ++LQK IEE
Sbjct: 266 LSKKHWSCKNPDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKGNKGIEE 325

Query: 304 HDKSWRENLAPOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQLGAQLRLTDNKGISRKIE 384

Query: 364 NRNYQFNYSLSLNPYTNLNLTAAYNSGRQKYPKAKFTGWLLKDFETYNNAKILDLNNT 423
NRNQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+
Sbjct: 385 NRNYQVNYNFNNNSYLDLNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

Query: 424 ATFRLPRETELQTTLGFNYFHNEYGKNRFPEELGLFFDGPQDQNGLYSY--LGRFGKDGK 481
TF LP+E +L+TLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K
Sbjct: 445 HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPEELSLFYNDASHDQGLYSHSKRGRYSGTGS 504

Query: 482 LLPQKSTIVQAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY
Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDLTALSKGIYHLNYSVNFTHYAFNGEYVGY----- 555

Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKKRANHSVSIADFGDYFMPFAGYSRTHRMP 601
EN+ + + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
Sbjct: 556 ---ENTAGQO-----INEPILHKSCHKKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI
Sbjct: 605 NIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIYI 664

Query: 662 HNVYKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAY 721
HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYIAHQNYKPIVKKSGVELEINYDMGRFFANVSYAY 722

Query: 722 QKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGRWLGNKLTLLGGA 781
Q++ QPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLLG A
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMPLPKDYGRLELGRWFDQKLTLLGLA 782

Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAEYEPKKN 841
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+
Sbjct: 783 ARYYGKSKRATIEEYINGSR-FKKNTLRRENYAVKKTEDIKKQPIILDHLVSYEPIKD 841

Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYYYSSFDPKDKDEDVTCNADKTLNKGKYGGS 901
LI +AEV+NL D+RY+DPLDAGNDAA+QRYYYSS + + C D + C GG+
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYYSS-----NNSIECAQDSSAC-----GGSD 892

Query: 902 KSVLTNFAARGRTFLMTMSYKF 922
K+VL NFARGRT+++++YKF
Sbjct: 893 KTVLYNFAARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
51  TTACGCGCTC CTGCGCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCGCGCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
15 401 CCGCCGCCAT CAACGGCAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTGC CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51  GYTALKMPAR AYELIPLAVL IGGLVSLSQL AAGSELTVIK ASGMSTKKLL
20 101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSVINVR EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACCTGA TTTCACGTTA CATCATCCGT CAAATGGCGG TTATGGCGGT
25 51  TTACGCGCTC CTGCGCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GAAATGCTG
151 GGCTACACCG CCCTCAAAAT GCCGCGCCGC GCCTACGAAC TGATTCCCCT
201 CGCCGTCCTT ATCGGCGGAC TGGTCTCCCT CAGCCAGCTT GCCGCGGCA
251 GCGAACTGAC CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CCGTCGCGCT
30 351 CGGCGAATGG GTTGCGCCCA CACTGAGCCA AAAAGCCGAA AACATCAAAG
401 CCGCCGCCAT CAACGGCAA ATCAGCACCG GCAATACCGG CCTTTGGCTG
451 AAAGAAAAAA ACAGCrTKAT CAATGTGCGC GAAATGTTGC CCGACCATAC
501 GCTTTTGGGC ATCAAAATT GGGCGCGCAA CGATAAAAC GAATTGGCAG
35 551 AGGCACTGGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGTTGGCAG
601 TTGAAAAACA TCCGCGCAG CACGCTTGGC GAAGACAAAG TCGAGGTCTC
651 TATTGCGGCT GAAGAAACT GGCCGATTTC CGTCAAACGC AACCTGATGG
701 ACGTATTGCT CGTCAAACCC GACCAAATGT CCGTCGGCGA ACTGACCACC
751 TACATCCGCC ACCTCCAAAA CAACAGCCAA AACACCCGAA TCTACGCCAT
801 CGCATGGTGG CGCAAATTGG TTTACCCCGC CGCAGCCTGG GTGATGGCGC
40 851 TCGTCGCCTT TGCCTTTACC CCGCAAACCA CCCGCCACGG CAATATGGGC
901 TTAAAACTCT TCGCGGCAT CTGTsTCGGA TTGCTGTTCC ACCTTGCCGG
951 ACGGCTCTTT GGGTTTACCA GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR QMAVMAVYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
45 51  GYTALKMPAR AYELIPLAVL IGGLVSLSQL AAGSELTVIK ASGMSTKKLL
101 LILSQFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151 KEKNSXINVR EMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
201 LKNIRRTSLG EDKVEVSIAA EENWPISVKR NLMDVLLVKP DQMSVGELTT
251 YIRHLQNNSQ NTRIIAIAWW RKLVPAAAW VMLVAFAFT PQTTTRHGNMG
50 301 LKLFGGICXG LLFHLAGRLF GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.meningitidis*:

5	orf112.pep	10 20 30 40 50 60	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
	orf112a	10 20 30 40 50 60	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
10	orf112.pep	70 80 90 100 110 120	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW
	orf112a	70 80 90 100 110 120	AYELMPLAVLIGGLVXSXQLAAGSELXVIKASGMSTKKLLLILSQFGFIFAIATVALGEW
15	orf112.pep	130 140 150 160	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH
	orf112a	130 140 150 160 170 180	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
20	orf112a	190 200 210 220 230 240	ELAEAVEADSAVLNSDGSWQLKNIRRLSTLGEDKVEVSIAAEEXWPISVKRNLMDVLLVKP

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCTT
30	201	CGCGTCTCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTTCG	TTTTATTTT	GCTATTGCCA	CCGTCGCGCT
	351	CGGCGAATGG	GTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
35	501	CCTGCTGGGC	ATTAAAATCT	GGGCCCGCAA	CGATAAAAAC	GAACCTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACGG	CAGTTGGCAG
	601	TTGAAAAACA	TCCGCCGCAG	CACGCTTGCC	GAAGACAAAG	TGCAGGTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTCGGCGA	ACTGACCACC
40	751	TACATCCGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTACGCCAT
	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTGCGCCT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
	901	TTAAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCTCTG
45	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAACACAG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

50	1	MNLISRYIIIR	QMAVMVAYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVXSXQL	AAGSELXVIK	ASGMSTKKLL
	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
	201	LKNIRRLSTL	EDKVEVSIAA	EEXWPISVKR	NLMDVLLVKP	DQMSVGELTT
	251	YIRHLQXXSQ	NTRIYAIWW	RKLVPAAAW	VMALVAFAPT	PQTTRHGNMG
55	301	LKXFGGICLG	LLFHLAAGRL	XFTSGLYGIP	PFLXGALPTI	AFALLAVWLI
	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

60	orf112a.pep	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR
	orf112-1	MNLISRYIIIRQMAVMVAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR
	orf112a.pep	AYELMPLAVLIGGLVXSXQLAAGSELXVIKASGMSTKKLLLILSQFGFIFAIATVALGEW

-485-

```

      orf112-1      ||||:||||||| |||||:||||||| |||||
5      orf112a.pep  VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN
      orf112-1      VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN
      orf112a.pep  ELAEAVEADSAVLNSDGSWQLKNIRSTLGEDKVEVSIAAEEXWPISVKRNLM DVLLVKP
10      orf112-1      ELAEAVEADSAVLNSDGSWQLKNIRSTLGEDKVEVSIAAEENWPISVKRNLM DVLLVKP
      orf112a.pep  DQMSVGELTTYIRHLQXXSQNTRIYAIAWWRKLVYPAAAWVMALVAFAPTQTTRHGNMG
      orf112-1      DQMSVGELTTYIRHLQNNSQNTRIYAIAWWRKLVYPAAAWVMALVAFAPTQTTRHGNMG
15      orf112a.pep  LKXFGGICLGLLFLHLAGRLFXFTS QLYGIPFLXGALPTIAFALLAVWLIRKQEKRX
      orf112-1      LKLFGGICXGLLFLHLAGRLFGFTSQL

```

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N. gonorrhoeae*:

```

25      orf112.pep  MNLISRYIIRQMAVMAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60
      orf112ng      MNLISRYIIRQMAVMAYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR 60
      orf112.pep  AYELIPLAVLIGGLVLSLSQLAAGSELTVIKASGMSTKKLLLILSQFGFIFAIATVALGEW 120
30      orf112ng      AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLILSQFGFIFAIAAVALGEW 120
      orf112.pep  VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH 166
      orf112ng      VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVREMLPDHTLLGIKIWARNDKN 180

```

The complete length ORF112ng nucleotide sequence <SEQ ID 891> is:

```

35      1 ATGAACCTGA TTTCACGTTA CATCATCCGC CAAATGGCGG TTATGGCGGT
      51 TTACGCGCTC CTTGCCTTCC TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
      101 ACGAAACCGG CAACCTCGGC AAAGGCAGTT ACGGCATATG GGAAATGCTG
      151 GGTACACCG CCTCAAAAT GCCCGCCCGC GCCTACGAAC TCATGCCCTC
40      201 CGCCGTCCCTC ATCGCGCGAC TGGCCTCTCT CAGCCAGCTT GCCGCCGGCA
      251 GCGAACTGCG CGTCATCAAA GCCAGCGGCA TGAGCACCAA AAAGCTGCTG
      301 TTGATTCTGT CTCAGTTCGG TTTTATTTT GCTATTGCCG CCGTCGCGCT
      351 CGGCGAATGG GTTCGCGCCA CGCTGAGCCA AAAAGCCGAA AACATCAAag
      401 cCGCCGCCAt taacggCAAA ATCAGCAccg gcAATACCGG CCTTTggcTG
45      451 AAAGAAAAAa ccAGCATTAT CAATGTGcGc GGAATGTTGC CCGACCATAC
      501 GCTTTTGGGC ATCAAAATT GGGCGCGCAA CGATAAAAC GAATTGGCAG
      551 AGGCAGTGA AGCCGATTCC GCCGTTTGA ACAGCGACGG CAGCTGGCAG
      601 TTGAAAAACA TCCGCCGCGC CATCATGGGT ACAGACAAAA TCGAAACATC
      651 cgCCGCGGCC GAAGAACTT gCGCGATTGC CGTCAGACGC AACCTGATGG
      701 ACGTATTGCT CGTCAAGCCC GACCAAATGT CCGTCGGCGA GCTGACCACC
50      751 TACATCCGCC ACCTCCAAA CAACAGCCAA AACACCCAAA TCTACGCCAT
      801 CGCATGGTGG CGTAAACTCG TTTACCCCGT CGCCGCATGG GTCATGGCGC
      851 TCGTTGCCTT CGCCTTTACG CCGCAAACCA CGCGCCACGG CAATATGGGC
      901 TTAATACTCT TCGCGGCAT CTGTCTCGGA TTGCTGTTCC ACCTTGCCGG
      951 CAGGCTCTTC GGGTTACCA GCCAACTCTA CGGCACCCCA CCCTTCCTCG
55      1001 CCGCGCACT GCCTACCATA GCCTTCGCCT TGCTCGCTGT TTGGCTGATA
      1051 CGCAAACAGG AAAACGTTG A

```

This encodes a protein having amino acid sequence <SEQ ID 892>:

```

60      1 MNLISRYIIR QMAVMAYAL LAFLALYSFF EILYETGNLG KGSYGIWEML
      51 GYTALKMPAR AYELMPLAVL IGGLASLSQL AAGSELAVIK ASGMSTKKLL
      101 LILSQFGFIF AIAAVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
      151 KEKTSIINVR GMLPDHTLLG IKIWARNDKN ELAEAVEADS AVLNSDGSWQ
      201 LKNIRRSIMG TDKIETSAAS EETWPIAVRR NLMDVLLVKP DQMSVGELTT
      251 YIRHLQNNQS NTQIYAIAW RKLVPVAAW VMALVAFAPT QTTRHGNMG
      301 LKLFGGICLG LLFHLAGRLE GFTS QLYGTP PFLAGALPTI AFALLAVWLI

```

351 RKQEK R*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLISRYIIRQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112-1	MNLISRYIIRQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
		10	20	30	40	50	60
10	orf112ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKLLLSQFGFIFAIAAVALGEW					
	orf112-1	AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSTKLLLSQFGFIFAIAATVALGEW					
		70	80	90	100	110	120
15	orf112ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINVRGMLPDHTLLGIKIWARNDKN					
	orf112-1	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSXINVREMLPDHTLLGIKIWARNDKN					
		130	140	150	160	170	180
20	orf112ng	ELAEAVEADSAVLNSDGSWQLKNIRRSIMGTDKIETSAAAEETWPIAVRRNLMVLLVKP					
	orf112-1	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEENWPISVKNRNLMDVLLVKP					
		190	200	210	220	230	240
25	orf112ng	DQMSVGELTTYIRHLQNNSONTOIYAIWWRLVYPVAAWVMALVAFATPQTRHGNMG					
	orf112-1	DQMSVGELTTYIRHLQNNSONTRIIYAIWWRLVYPAAWVMALVAFATPQTRHGNMG					
		250	260	270	280	290	300
30	orf112ng	LKLFGGICLGLLFLHLAGRLFGFTSQLYGTPPFLAGALPTIAFALLAVWLIRKQEKRX					
	orf112-1	LKLFGGICXGLLFLHLAGRLFGFTSQL					
		310	320	330	340	350	
35	orf112ng						
	orf112-1						

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	<u>CGCGGATCCGCTAGC</u> -GGACACACTTATTTCTGG	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	<u>GCGGATCCCATATG</u> -TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	<u>GCGGATCCCATATG</u> -TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGATTTACGGACGCGCA	
ORF 4	Forward	<u>GCGGATCCCATATG</u> -TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAACC	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-ATGGAAGGCGCACAAC	
	Reverse	CCCGCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	<u>CGCGGATCCCATATG</u> -ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	<u>CGCGGATCCGCTAGC</u> -GCGCTGCTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-TTCAAATATATTTGCGGA	
ORF 8	Forward	<u>GCGGATCCCATATG</u> -GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	<u>CGCGGATCCCATATG</u> -CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCCGAGGTTTTCGGG	
ORF 10	Forward	<u>GCGGATCCCATATG</u> -GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG- TAATGGGAAACCTTGTTTT	
ORF 11	Forward	<u>GCGGATCCCATATG</u> -GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	<u>CGCGGATCCCATATG</u> -GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGGGTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TGCGGGACACTGACAGG	
	Reverse	CCCGCTCGAG-AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTTCG	NdeI-NcoI

	Forward	CGGGATCC-ATTGCCGGCCTGTTCG	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTTGTACAGC	XhoI
ORF 18	Forward	GCGGATCCCATATG-ATTTTGCTGCATTTGGAT	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	XhoI
ORF 19	Forward	GGAATTCCATATGGCCATGG -TCGCCAGTGTTTTACC	NdeI-NcoI
	Reverse	CGGGATCC-TTCGCCAGTGTTTTACCG	BamHI
	Forward	CCCGCTCGAG-GGTGTTTTTGAAGCTGCC	XhoI
ORF 20	Forward	GGAATTCCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Reverse	CGGGATCC-TTCGGCGCGGGTATG	BamHI
	Forward	CCCGCTCGAG-CGGCGAGCGAGAGCA	XhoI
ORF 22	Forward	GGAATTCCATATGGCCATGG-TGATTAAAATCAAAAAGGTCT	NdeI-NcoI
	Reverse	CGGGATCC-ATGATTAAAATCAAAAAGGTCTAAACC	BamHI
	Forward	CCCGCTCGAG-ATTATGATAGCGGCC	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG	XhoI
ORF 24	Forward	GGAATTCCATATGGCCATGG -TGATGCCGGAATGGTG	NdeI-NcoI
	Reverse	CGGGATCC-ATGATGCCGGAATGGTG	BamHI
	Forward	CCCGCTCGAG-TGTCAGCGTGGCGCA	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGCGTTTT	XhoI
ORF 27	Forward	GGAATTCCATATGGCCATGG-AGACCTATTCTGTTTA	NdeI-NcoI
	Reverse	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC	BamHI
	Forward	CCCGCTCGAG-GGGTTCGATTAAATAACCAT	XhoI
ORF 28	Forward	GGAATTCCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
	Reverse	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
	Forward	CCCGCTCGAG-TTTGTCTAGAGGAATTCGCG	XhoI
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGATGCCCCG	BamHI-NdeI
	Reverse	CGCGGATCCGCTAGC-AACGGTTTGATGCCCCG	BamHI-NheI
	Forward	CCCGCTCGAG-TTTGTCTAAGTTCTGATATG	XhoI
ORF 32	Forward	CGCGGATCCCATATG-AATACTCTCTTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTTTGATGCTTTG	XhoI
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC	XhoI

ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT	BamHI-NdeI BamHI-NheI XhoI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT	
	Reverse	CCCGCTCGAG-AAACAGCCATTTGAGCGA	
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATAGCCCCTTTCAGG	
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGGAC	
ORF 65	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG	
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA	
ORF 66	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGGATTTTGCAGAGATGG	
ORF 72	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA	
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTTCG	
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG	
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC	NheI-NdeI BamHI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT	
ORF 79	Forward	CGCGGATCCCATATG-GTTTCGCGCCCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG	
ORF 83	Forward	GCGGATCCCATATG-AAACCCCTGCTGCTGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCGCCTTTGCGGC	
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA	
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TCGGCGCGGCGGGC	
ORF 89	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG	
	Reverse	CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC	
ORF 97	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTTCCGGCAAATC	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCCGCCTTTCAATGT	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTCCCGATGATGTT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCAACCGAAG	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAGGCAAAAACCGCA	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGTTTCGGGCAGGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAAGACGG	EcoRI
	Reverse	AAAAAAGTCGAC-CTATTTTTAGGGGCTTTTGC TTGTTGAAAAGCCTGCC	Sall
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTGCC	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCCTGGCCGTTGAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	PstI

ORF122	Forward Reverse	AAAAAAGTCGAC-ATGCTTACCGCGCAAGCAGTTCTCC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
ORF125	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
ORF126	Forward Reverse	AAAGAATTC-GCGGAAACGGTCAAG AAACTGCAG-TTAATCTGTCTTCCGATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTTGCACTCCTTATC	EcoRI Sall
ORF128	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGCTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTGAGCAGGCG	EcoRI PstI
ORF129	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTGTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTGGGGCGG	EcoRI PstI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATTCCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
ORF 131	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward Reverse	GCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTTGTAGCTCATCGT	BamHI-NdeI XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward Reverse	GCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACTTT	BamHI-NheI XhoI
ORF 137	Forward Reverse	GCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAAACAGC	BamHI-NdeI XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTAAATATT	BamHI-NdeI XhoI
ORF 142	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
ORF 143	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
ORF 144	Forward Reverse	GCGGATCCCATATG-ACCTTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
ORF 147	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..

7. An antibody which binds to a protein according to any one of claims 4 to 6.

8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.

9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
 371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
 411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
 5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
 491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
 531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
 571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
 611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
 10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
 691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
 731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
 15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
 & 891..

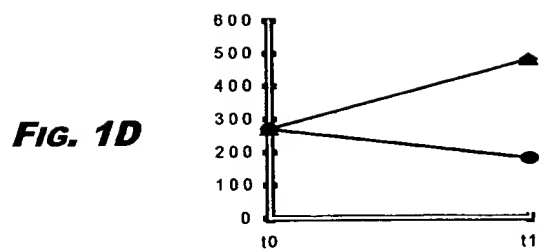
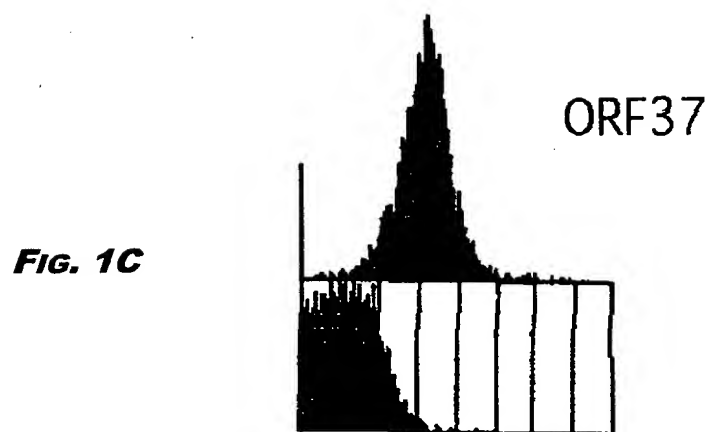
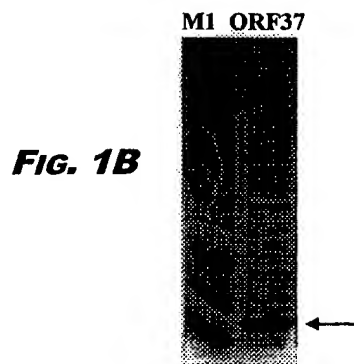
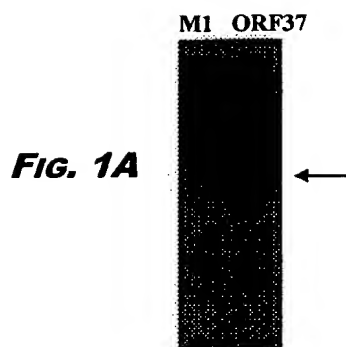
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
 group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
 20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
 25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
 30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
 35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

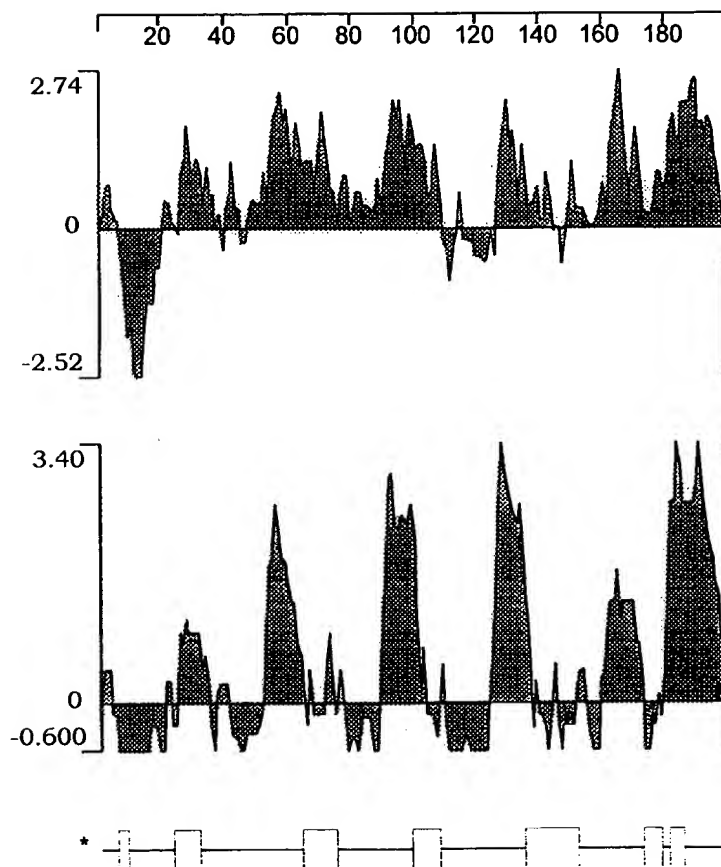
- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

FIGURE 1



2/24

FIG 1E

3/24

FIGURE 2

FIG. 2A

M1 ORF5

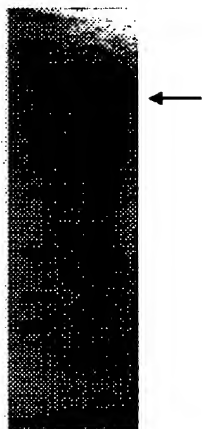
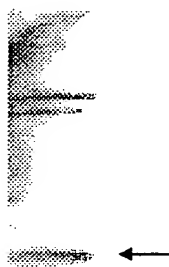
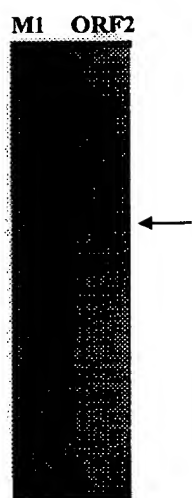
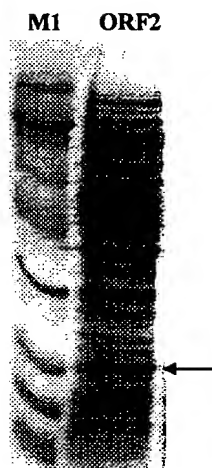
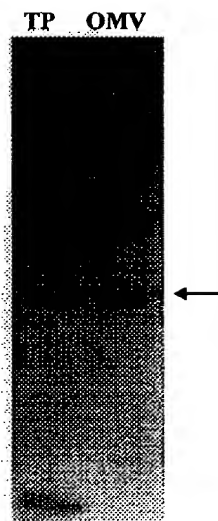
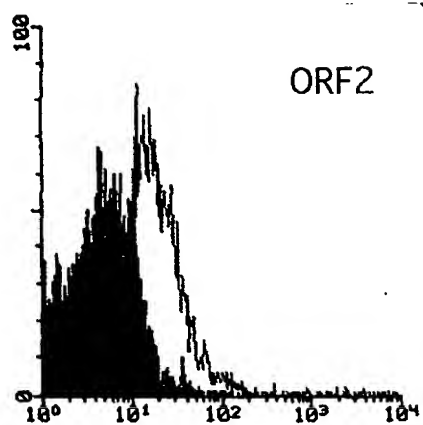


FIG. 2B

TP



4/24

FIGURE 3**FIG. 3A****FIG. 3B****FIG. 3C****FIG. 3D**

5/24

FIGURE 4

FIG. 4A

M1 ORF15

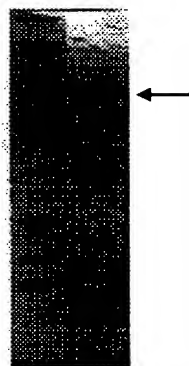


FIG. 4B

M2 ORF15

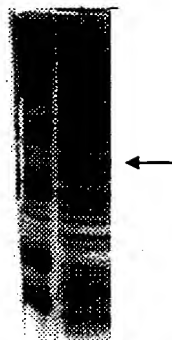
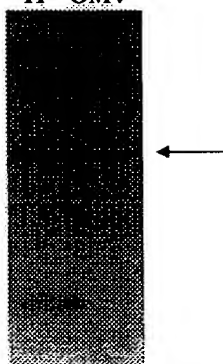


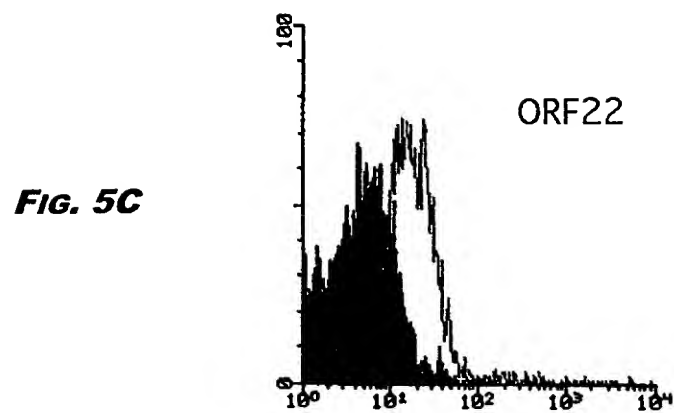
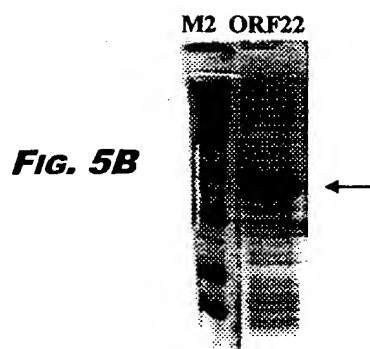
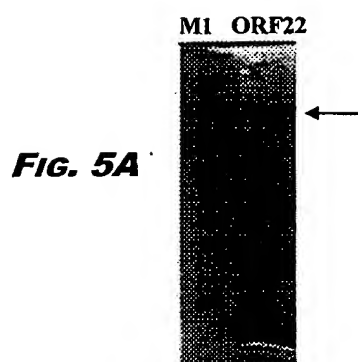
FIG 4C

TP OMV



6/24

FIGURE 5



7/24

FIGURE 6

Fig. 6A

M1 ORF28

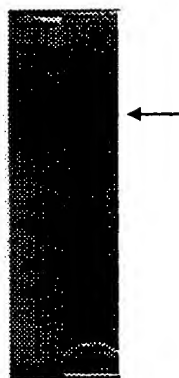


Fig. 6B

M2 ORF28

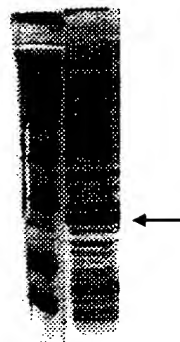


FIGURE 7

Fig. 7A

M1 ORF32

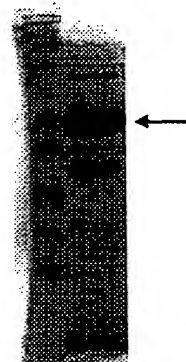
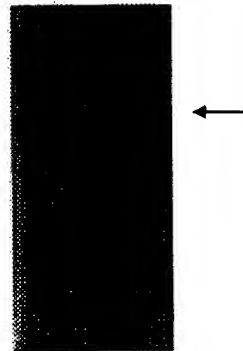


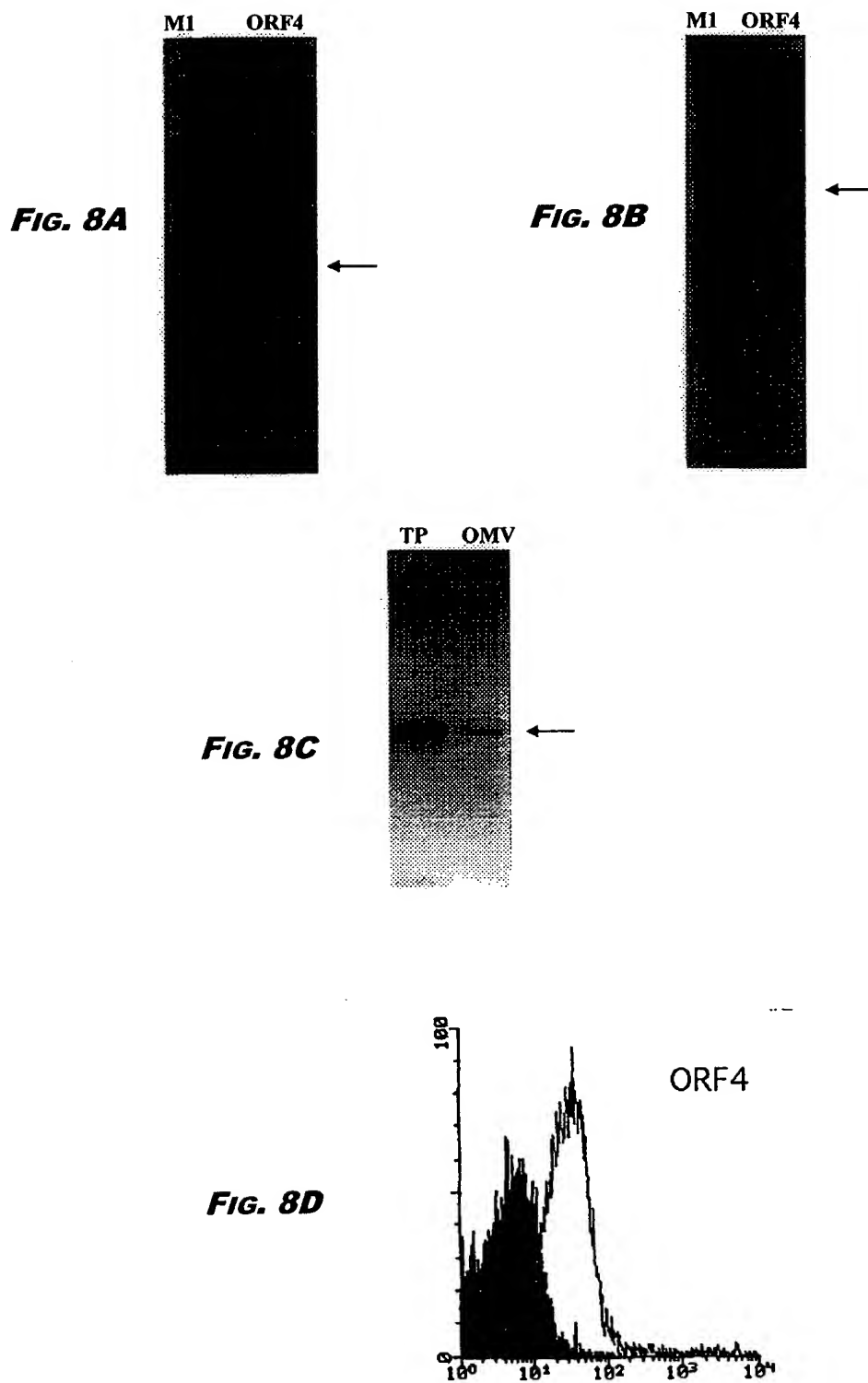
Fig. 7B

M1 ORF32

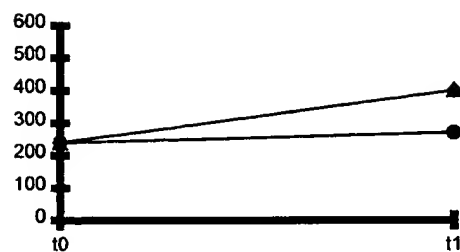
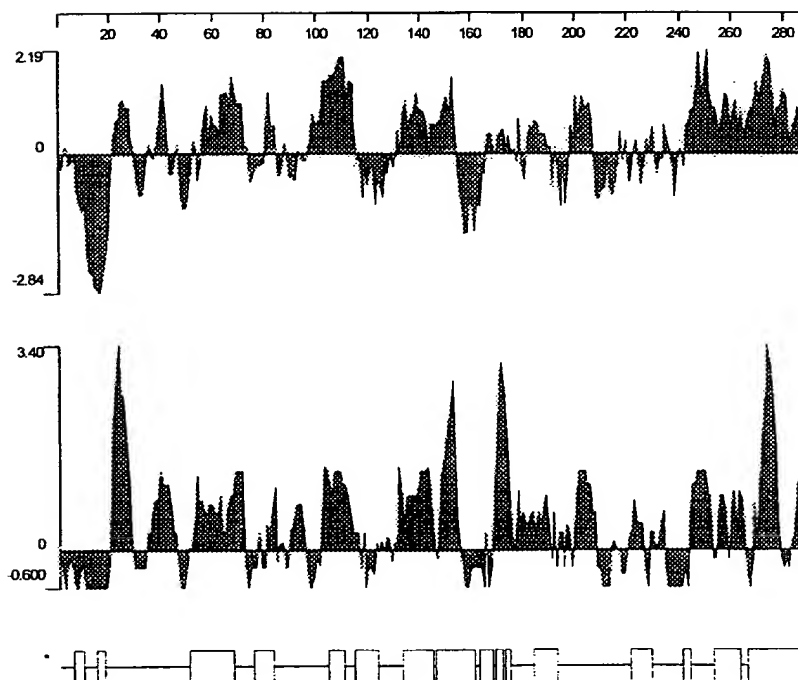


8/24

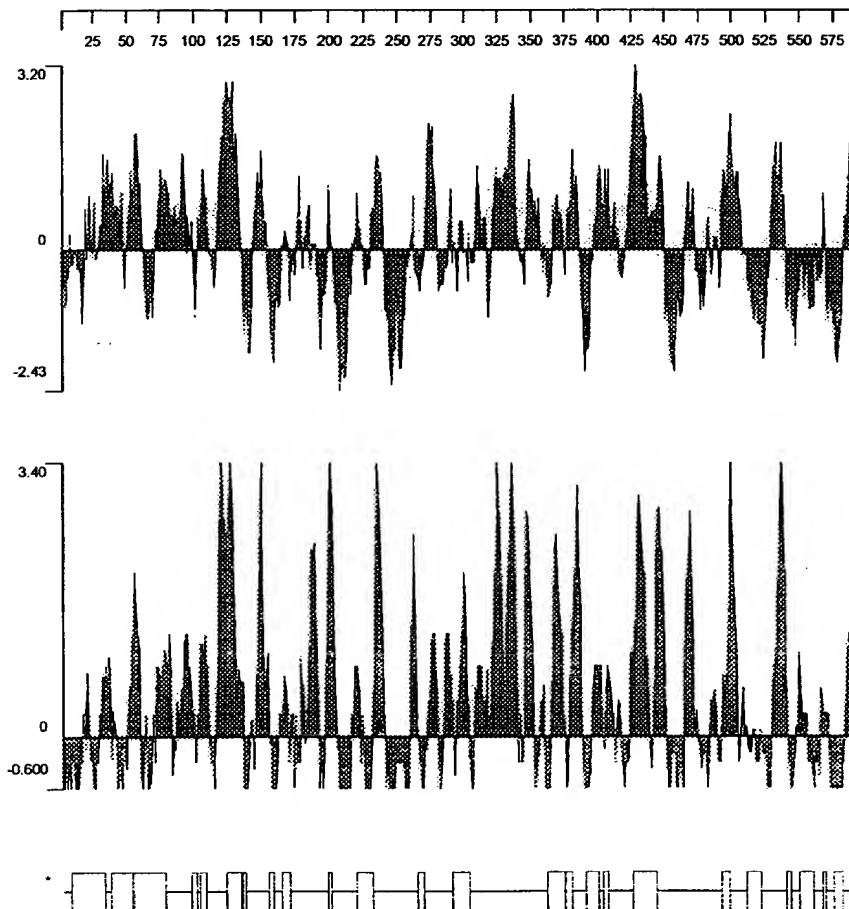
FIGURE 8



9/24

Fig. 8E**Fig. 8F**

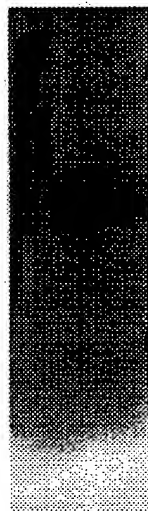
10/24

FIGURE 9

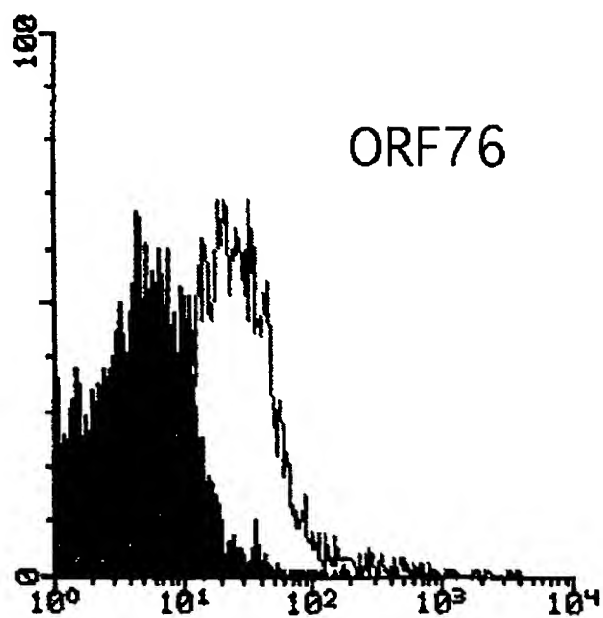
11/24

FIGURE 10**Fig. 10A**

M1 ORF76

**Fig. 10B**

TP OMV

**Fig. 10C**

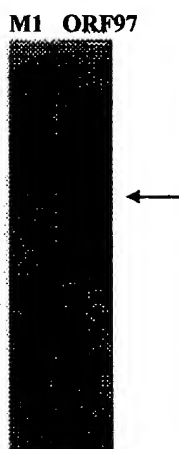
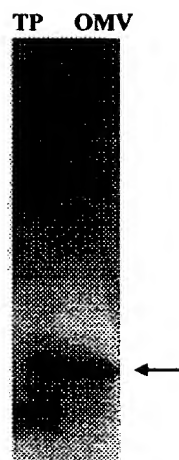
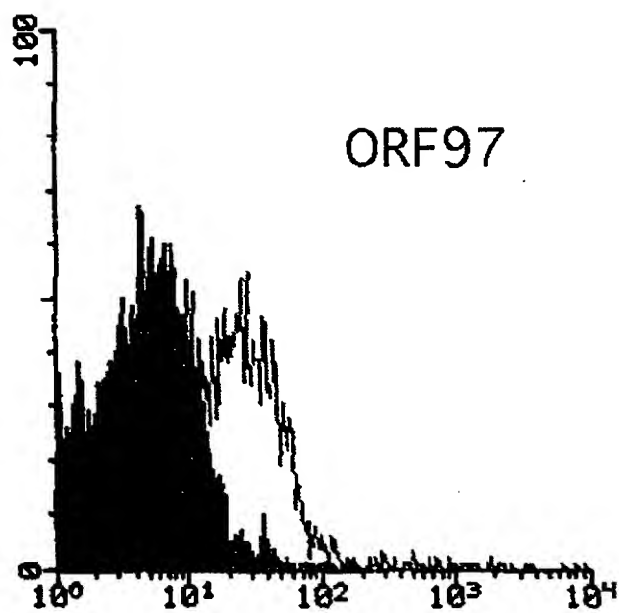
12/24

FIGURE 11

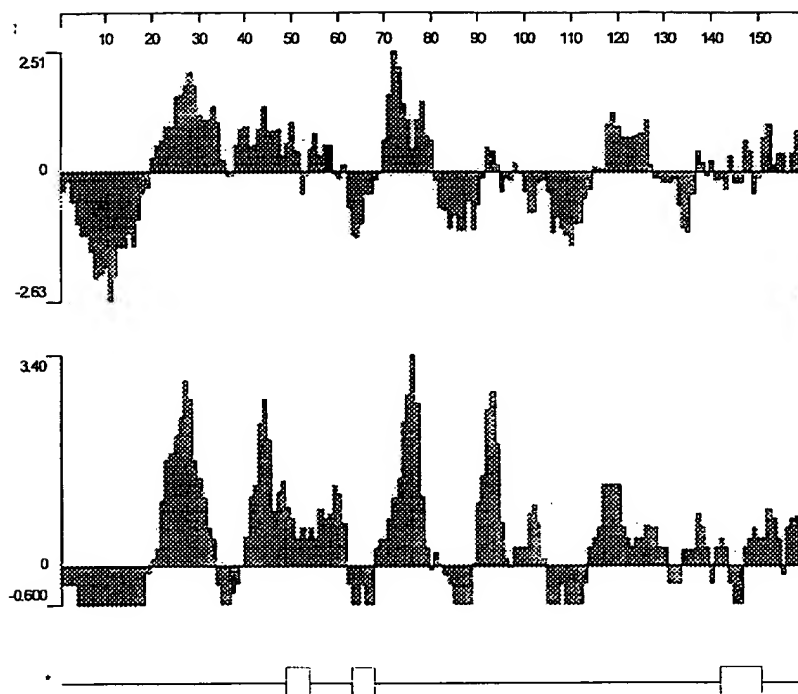
M1 ORF89



13/24

FIGURE 12**Fig. 12A****Fig. 12B****Fig. 12C****Fig. 12D**

14/24

FIG. 12E

15/24

FIGURE 13

Fig. 13A

M1 ORF106

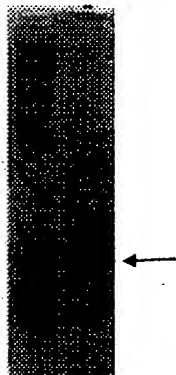


Fig. 13B

M2 ORF106

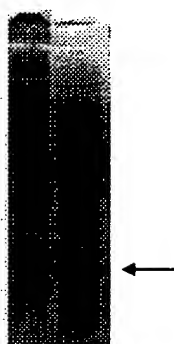
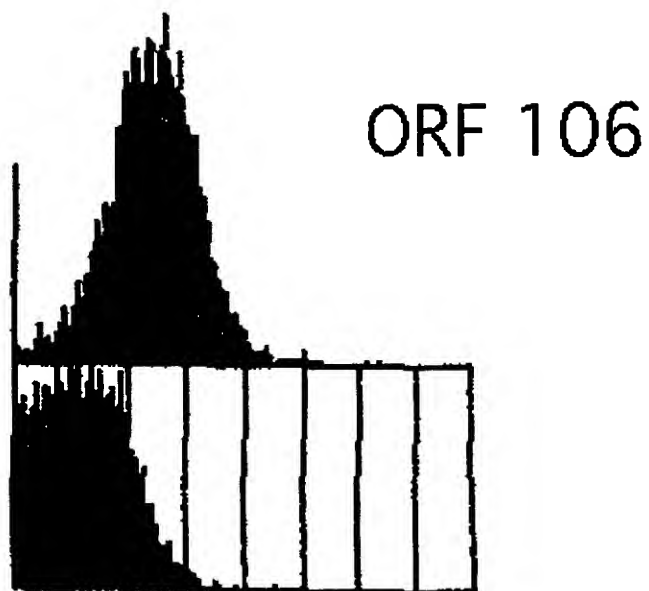
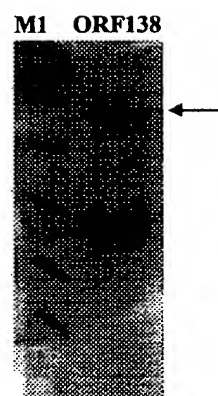
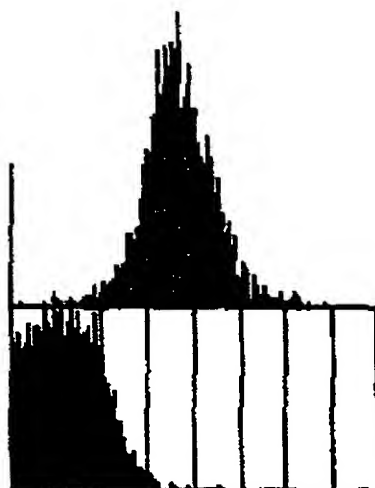


Fig. 13C



16/24

FIGURE 14**FIG. 14A****FIG. 14B**

17/24

FIGURE 15

Fig. 15A

M1 ORF23

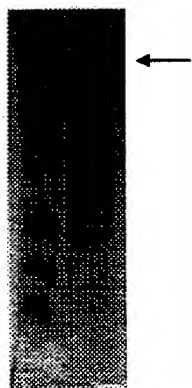


Fig. 15B

M2 ORF23

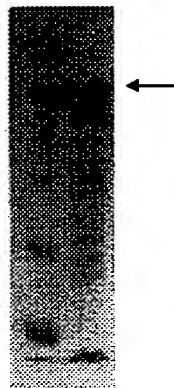
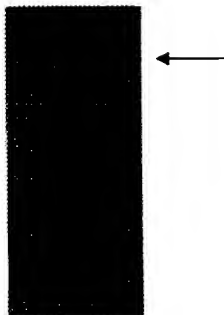
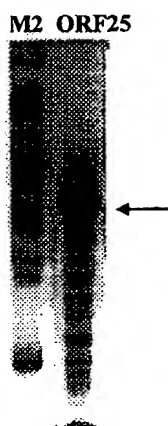
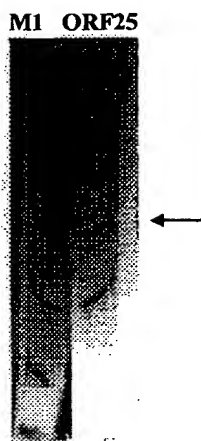
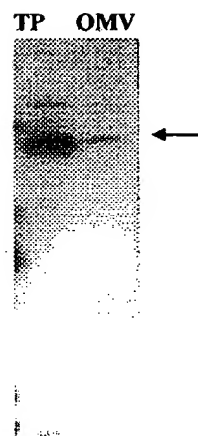
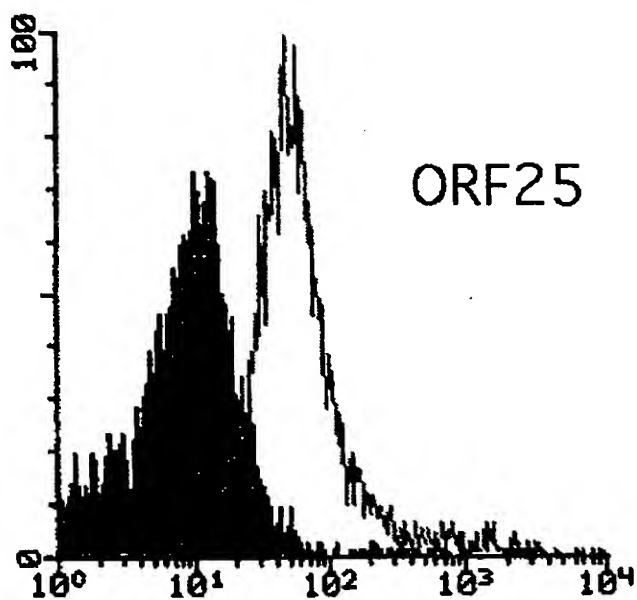


Fig 15C

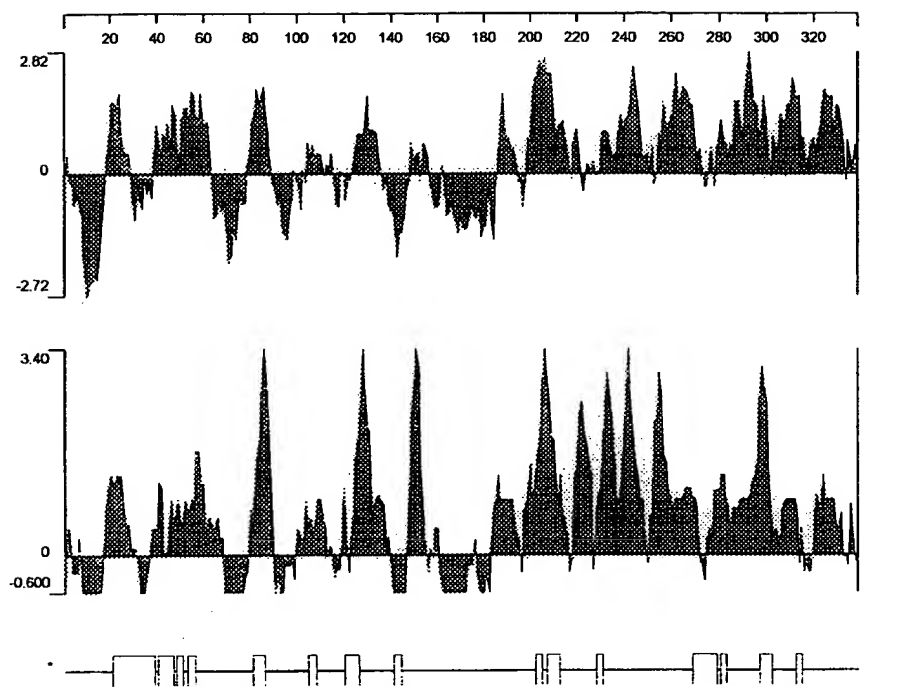
TP OMV



18/24

FIGURE 16**Fig. 16A****Fig. 16B****Fig. 16C****Fig. 16D**

19/24

FIG. 16E

20/24

FIGURE 17

FIG. 17A

M1 ORF27

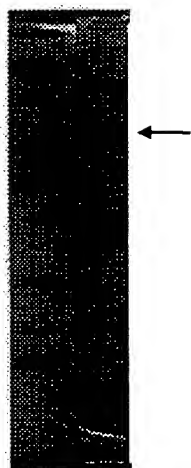
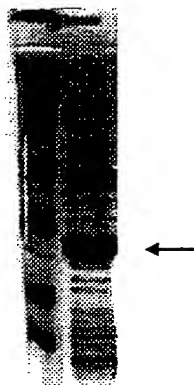
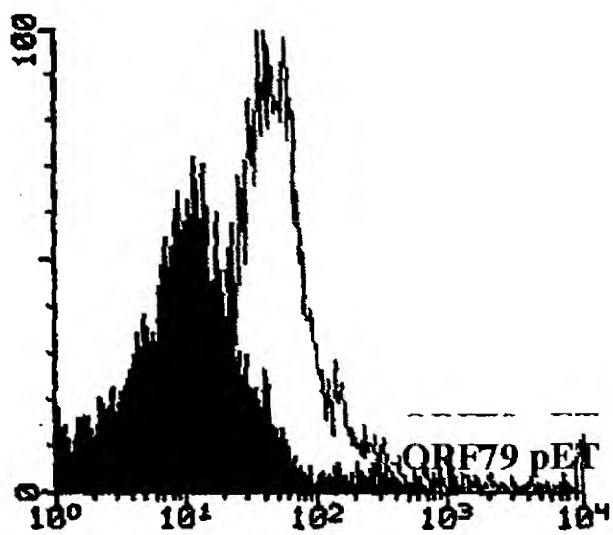


FIG. 17B

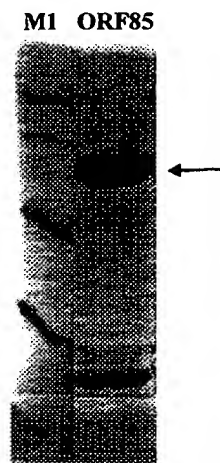
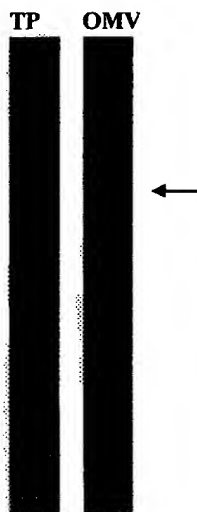
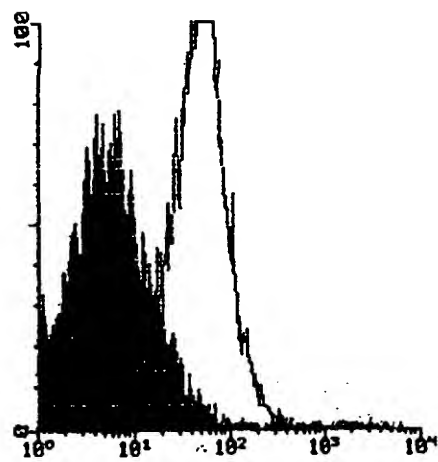
M2 ORF27



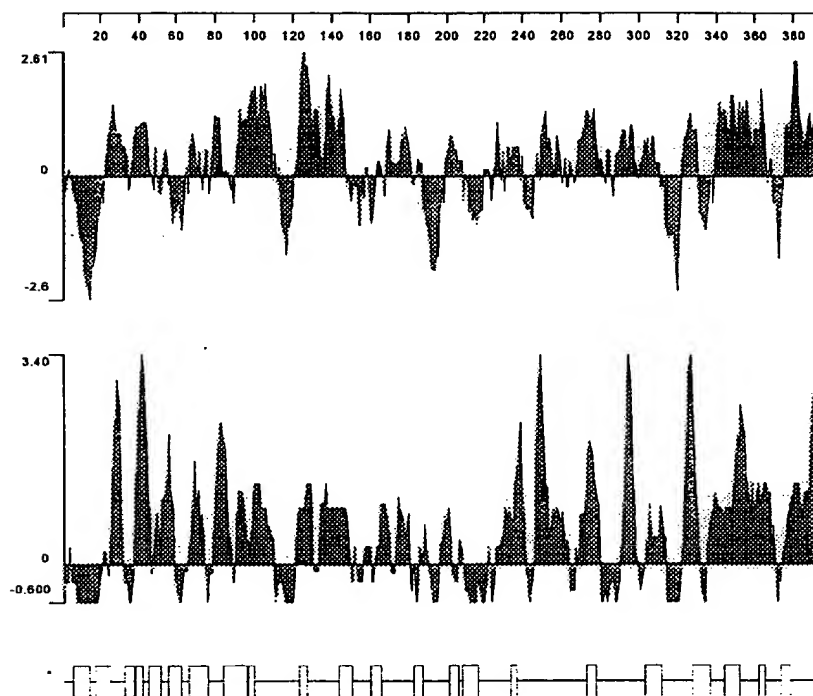
21/24

FIGURE 18**FIG. 18A****FIG. 18B**

22/24

FIGURE 19**Fig. 19A****Fig. 19B****Fig. 19C**

23/24

Fig 19D

24/24

FIGURE 20

Fig. 20A

M1 ORF132



Fig. 20B

M2 ORF132



Fig. 20C



ORF132